



BACHELORARBEIT

Herr
André Marquardt

**Entwicklung eines
Vorhersagealgorithmus für
Entfaltungsbarrieren in
 α -helikalen Membranproteinen
und Anwendung auf Mutationen
und homologe Proteine**

2013

BACHELORARBEIT

Entwicklung eines Vorhersagealgorithmus für Entfaltungsbarrieren in α -helikalen Membranproteinen und Anwendung auf Mutationen und homologe Proteine

Autor:

André Marquardt

Studiengang:

Biotechnologie/Bioinformatik

Seminargruppe:

Bi10w2-B

Erstprüfer:

Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer:

M.Sc. Florian Heinke

Mittweida, August 2013

Danksagung

An dieser Stelle möchte ich die Möglichkeit nutzen, einigen Personen meinen ausdrücklichen Dank auszusprechen. An erster Stelle danke ich Prof. Dr. Dirk Labudde, dessen Idee das Thema dieser Arbeit war. Weiterhin möchte ich mich für seine langjährige Betreuung und Motivation bedanken. Ein weiterer großer Dank gilt meinem Betreuer M.Sc. Florian Heinke. Er war nicht nur Mentor sondern auch ständiger Ideengeber und Motivator. Natürlich gilt auch meinen Eltern ein gebührender Dank ohne deren Unterstützung, sowohl zur Schulzeit als auch im Studium, ich niemals so weit gekommen wäre.

I. Inhaltsverzeichnis

Inhaltsverzeichnis	I
Abbildungsverzeichnis	II
Tabellenverzeichnis	III
Vorwort	IV
1 Einleitung	1
1.1 Proteine: Aufbau und Funktion	1
1.2 3D-Struktur und Strukturaufklärung	2
1.2.1 Röntgenstrukturanalyse	2
1.2.2 Kernspinresonanzspektroskopie (NMR)	2
1.3 Theoretische und Praktische Methoden	3
1.3.1 Anisotropic Network Model (ANM)	3
1.3.2 Energieprofil und Kontaktprofil	4
1.3.3 3D-Strukturvorhersage	5
1.3.4 Single Molecule Force Spectroscopy (SMFS)	6
1.3.5 In silico-Simulation des SMFS-Entfaltens	8
1.4 Verwendete Proteine	8
1.4.1 Bacteriorhodopsin	9
1.4.2 Halorhodopsin	10
1.4.3 Humanes Aquaporin 1	11
1.4.4 Natives Rinderrhodopsin	12
1.4.5 Mutationen	13
1.5 Gütemaße	13
1.5.1 Konfusionsmatrix	13
1.5.2 Matthews Correlation Coefficient (MCC)	14
1.5.3 F1-Maß (F1)	14
2 Barrieren-Vorhersagealgorithmus	15
2.1 Pseudocode	15
2.2 Parametrisierung	18
3 Ergebnisse	21
3.1 Bacteriorhodopsin	22
3.1.1 Mutationen	24
3.1.2 Mutante - D96N	25
3.1.3 Mutante - D85S	26
3.1.4 Mutante - F219L	27
3.1.5 Mutante - P50A	28
3.1.6 Mutante - M56A	29
3.1.7 Mutante - P186A	30
3.1.8 Mutante - P91A	31

3.1.9	Mutante - T24S	32
3.1.10	Mutante - T24V	33
3.1.11	Mutante - T24A	34
3.1.12	Mutante - G113L	35
3.1.13	Mutante - G113Q	36
3.1.14	Mutante - G116L	37
3.1.15	Mutante - G116Q	38
3.1.16	Mutante - I117A	39
3.1.17	Mutante - I117F	40
3.1.18	Mutante - M145F	41
3.2	Halorhodopsin	42
3.2.1	Mutationen	44
3.2.2	Mutante - H95A	45
3.2.3	Mutante - H95R	46
3.2.4	Mutante - R108Q	47
3.2.5	Mutante - T203V	48
3.3	Humane Aquaporine 1	49
3.3.1	Vergleich der aufgeklärten humanen Aquaporine 1	50
3.3.2	Vergleich der vorhergesagten Barrieren von Aquaporin 1 bis 11	53
3.4	Natives Rinderrhodopsin	56
3.5	Zusammenfassung der Ergebnisse	58
4	Evaluation	59
4.1	Mit Experiment-Simulation	59
4.2	Nativer Zustand	60
4.3	Signifikanztest mittels randomisierter Daten	61
5	Diskussion	65
5.1	Mutationen	65
5.2	Vergleich Bacterio- und Halorhodopsin	72
5.3	Vergleich humane Aquaporine 1	74
5.4	Vergleich Aquaporine 1 bis 11	77
5.5	Vergleich Vorhersage mit randomisierten Daten	79
6	Molekularbiologische Interpretation	81
7	Zusammenfassung	83
8	Ausblick	85
	Literaturverzeichnis	89

II. Abbildungsverzeichnis

1.1 ANM Trajektorie	3
1.2 Erklärung SMFS-Vorgang	6
1.3 Beispielhaftes SMFS-Ergebnis	7
1.4 Idealisierter SMFS-Vorgang	7
1.5 Zusammenfassung SMFS-Experiment	8
1.6 Bacteriorhodopsin	9
1.7 Halorhodopsin	10
1.8 Aquaporin 1	11
1.9 Rinderrhodopsin	12
2.1 Funktionsweise Algorithmus Schritt 1	15
2.2 Funktionsweise Algorithmus Schritt 2	16
2.3 Funktionsweise Algorithmus Schritt 3	17
2.4 Funktionsweise Algorithmus Schritt 4	18
3.1 Bildliche Ergebnisdarstellung der Vorhersage von Bacteriorhodopsin	23
3.2 Bildliche Ergebnisdarstellung der Vorhersage von Halorhodopsin	43
3.3 Bildliche Ergebnisdarstellung der Vorhersage von Aquaporin 1	51
3.4 Bildliche Ergebnisdarstellung der Vorhersage von Rinderrhodopsin	57
3.5 Schematischer Vergleich alle Proteine	58
5.1 Energieprofilvergleich Bacteriorhodopsin zu einer Mutation	68
5.2 Kontaktprofilvergleich Bacteriorhodopsin zu einer Mutation	69
5.3 Energieprofilvergleich Halorhodopsin zu einer Mutation	70
5.4 Kontaktprofilvergleich Halorhodopsin zu einer Mutation	71
5.5 Vergleich Bacterio- und Halorhodopsin	73
5.6 Energieprofile Aquaporine 1	75
5.7 Kontaktprofile humane Aquaporine 1	76
5.8 Grafischer Vergleich alle Aquaporine	78
8.1 Vergleich Archae- Bacterio- Sensoryrhodopsin	85
8.2 Vergleich Archaeerhodopsin 1 und 2 und ASR und BR	86

III. Tabellenverzeichnis

1.1 Allgemeine Konfusionsmatrix [KP98]	13
3.1 Ergebnisdarstellung für Bacteriorhodopsin.	22
3.2 Ergebnisdarstellung für den D96N Mutanten 1C8R.	25
3.3 Ergebnisdarstellung für den D85S Mutanten 1JV7.	26
3.4 Ergebnisdarstellung für den F219L Mutanten 1JV7.	27
3.5 Ergebnisdarstellung für den P50A Mutanten 1PXR.	28
3.6 Ergebnisdarstellung für den M56A Mutanten 1JV7.	29
3.7 Ergebnisdarstellung für den P186A Mutanten 1Q5I.	30
3.8 Ergebnisdarstellung für den P91A Mutanten 1Q5J.	31
3.9 Ergebnisdarstellung für den T24S Mutanten 1S51.	32
3.10 Ergebnisdarstellung für den T24V Mutanten 1S52.	33
3.11 Ergebnisdarstellung für den T24A Mutanten 1S54.	34
3.12 Ergebnisdarstellung für den G113L Mutanten von Bacteriorhodopsin.	35
3.13 Ergebnisdarstellung für den G113Q Mutanten von Bacteriorhodopsin.	36
3.14 Ergebnisdarstellung für den G116L Mutanten von Bacteriorhodopsin.	37
3.15 Ergebnisdarstellung für den G116Q Mutanten von Bacteriorhodopsin.	38
3.16 Ergebnisdarstellung für den G117A Mutanten von Bacteriorhodopsin.	39
3.17 Ergebnisdarstellung für den I117F Mutanten von Bacteriorhodopsin.	40
3.18 Ergebnisdarstellung für den M145F Mutanten von Bacteriorhodopsin.	41
3.19 Ergebnisdarstellung für Halorhodopsin.	42
3.20 Ergebnisdarstellung für den H95A Mutanten von Halorhodopsin.	45
3.21 Ergebnisdarstellung für den H95A Mutanten von Halorhodopsin.	46
3.22 Ergebnisdarstellung für den R108Q Mutanten von Halorhodopsin.	47
3.23 Ergebnisdarstellung für den T203V Mutanten von Halorhodopsin.	48
3.24 Ergebnisdarstellung für die drei unterschiedlich aufgeklärten humanen Aquaporine 1.	50
3.25 Ergebnisdarstellung für die Aquaporine 1 bis 6.	54
3.26 Ergebnisdarstellung für die Aquaporine 7 bis 11.	55
3.27 Ergebnisdarstellung für natives Rinderrhodopsin.	56

4.1 Darstellung der Ergebnisse für die Simulation des Experiments	59
4.2 Konfusionsmatrix für die Simulation des Experiments	59
4.3 Darstellung der Ergebnisse für den nativen Zustand	60
4.4 Konfusionsmatrix für die Simulation des Experiments	61
4.5 Aufeinanderfolgen von Vorhersagen	62
4.6 Wahrscheinlichkeit der Abfolge von Vorhersagen	62
5.1 Einfluss von Mutationen auf die Struktur	66

Bibliografische Angaben

Marquardt, André: Entwicklung eines Vorhersagealgorithmus für Entfaltungsbarrieren in α -helikalen Membranproteinen und Anwendung auf Mutationen und homologe Proteine, 95 Seiten, 28 Abbildungen, 35 Tabellen, Hochschule Mittweida (FH) - University of Applied Sciences, Fakultät Mathematik / Naturwissenschaften / Informatik

Bachelorarbeit, 2013

Zusammenfassung

Die Wichtigkeit von Proteinen für das Leben von Organismen ist seit ihrer Entdeckung bekannt. Allgemein werden Proteine in zwei Hauptgruppen eingeteilt, in globuläre und Membranproteine. Globuläre Proteine kommen meistens in wässriger Umgebung vor (z. B. im Cytoplasma oder in extrazellulären, von der Zelle produzierten, Flüssigkeiten), wohingegen Membranproteine entweder an der Membran aufgebracht sind oder sie durchdringen und deshalb hydrophob molekularen Kräften ausgesetzt sind, welche durch die Lipide und andere hydrophobe Makromoleküle verursacht werden. Dadurch, dass sie als Flaschenhalsmoleküle, in praktisch unzähligen biologischen Prozessen, agieren, wie zum Beispiel passiver und aktiver Molekültransport und Signalvermittlung zwischen Zellen und Zellorganellen, ermöglichen Membranproteine das Leben mit seiner bekannten Komplexität. Deshalb ist das Verstehen der Mechanismen von Membranproteinen der Schlüssel zum Begreifen von zellulären Prozessen und vermag erheblich zur Grundlage der next-generation Diagnostik beizutragen. Um diese Mechanismen erklären zu können ist es eine Notwendigkeit die Beziehung zwischen der atomaren Proteinstruktur und der zugrundeliegenden molekularen Dynamik, welche die Proteinfunktion regeln, zu analysieren. Dennoch ist es, aufgrund ihrer hydrophoben physikochemischen Eigenschaften ein herausfordernder Prozess, Erkenntnisse über die Dynamik von Membranproteinen zu gewinnen, da andere experimentelle Vorgehen benötigt werden, als sie sonst bei globulären Proteinen angewendet werden. Single Molecule Force Spectroscopy (SMFS) hat sich selbst als verlässlicher Ansatz zur direkten Messung von intra- und intermolekularen Kräften, welche in α -helikalen Membranproteinen vorhanden sind, bewiesen. Im Wesentlichen kann SMFS dazu benutzt werden, Proteine kontinuierlich und kontrolliert zu entfalten und dabei das Entfaltungsgeschehen, Entfaltungsbarrieren und die benötigte Kräfte um diese aufzubrechen, zu verfolgen und Aufschluss über stabilisierende Eigenschaften des Proteins bei der Untersuchung zu geben. Allerdings sind die zugrundeliegenden theoretischen Aspekte und Details, wie sich ein Protein während des SMFS-Experiments entfaltet, noch nicht komplett verstanden.

In der Bemühung, die beim SMFS-Experiment gemachten Beobachtungen zu unterstützen und die experimentelle Dynamik auf atomarer Ebene zu bestärken, wird in dieser Arbeit ein Modell zur Vorhersage von Entfaltungsbarrieren und stabilisierenden Regionen in α -helikalen Membranproteinen vorgestellt und evaluiert. Außerdem wird der entwickelte Vorhersagealgorithmus sowohl auf den Wildtyp von Proteinen, als auch auf Mutanten, welche in der Literatur gut beschrieben sind, angewendet. Die Ergebnisse zeigen den globalen Effekt der Mutationen auf die Stabilität des Proteins, verglichen zum Wildtyp. Hier wird der globale Einfluss der Mutationen auf die Stabilität von Proteinen durch den Vergleich der zugrundeliegenden Vorhersageeigenschaften, welche vom Wildtyp und den mutierten Proteinen abgeleitet werden, gezeigt und evaluiert.

Schlussendlich führen erste Studien bei der Vorhersage von Entfaltungsbarrieren bei homologen Membranproteinen zu der Hypothese, dass Entfaltungsbarrieren und stabilisierende physikochemische Eigenschaften genauso konserviert sind wie die gemeinsame Faltung des Proteins.

Abstract

The importance of proteins for life of organisms is known since their discovery. In general, proteins are classified into two main groups, globular and membrane proteins. Globular proteins are mostly present in aqueous environments (f. e. the cytoplasm or extra-cellular fluids produced by the cell), whereas membrane proteins are attached or integrated into the membrane (bi-)layer and hence are exposed to hydrophobic molecular forces caused by lipids and other hydrophobic macromolecules.

Acting as bottle-neck molecules in virtually countless biological processes, such as passive and active molecule transport and signal mediation between cells or cell organelles, membrane proteins afford life in its known complexity. Thus, understanding membrane protein mechanisms is the key for comprehending cellular processes and can contribute significantly to the basis of next-generation diagnostics. To explain these mechanisms it is a necessity to analyse the relationship between the atomic protein structure and the underlying molecular dynamics that govern protein functionality. However, due to their hydrophobic physiochemical properties gaining knowledge on membrane protein dynamics is a challenging process and requires much different experimental approaches that are usually applied in the analyses of globular proteins. Single molecule force spectroscopy (SMFS) has proven itself to be a valuable approach to directly measure intra- and intermolecular forces present in α -helical membrane proteins. In essence, SMFS can be used to unfold proteins continuously in a controlled manner and trace unfolding events, energetic barriers and required forces to break them - and finally give conclusions about stabilising characteristics of the protein under investigation. However, the underlying theoretical aspects and details on how proteins unfold in SMFS experiments are not yet understood entirely. In an effort to support findings obtained using SMFS and fortify the experimental dynamics on atomic level, a model for predicting unfolding barriers and stabilising regions in α -helical membrane proteins is presented and evaluated in this work. Furthermore, the developed prediction algorithm is applied to wild-type proteins and mutants which are well described in the literature. Results demonstrate the global effect of mutations on stability of proteins, compared to the wild type. Here, the global influence of mutations on stability of proteins is shown and evaluated by the comparison of the underlying prediction features derived from wild-type and mutant proteins. Finally, first studies of predicted unfolding barriers in homologous membrane proteins led to the hypothesis that unfolding barriers and stabilising physiochemical characteristics are as conserved as the common protein fold.

IV. Vorwort

If the facts don't fit the theory, change the facts.

Albert Einstein

Proteine, unter ihnen die Membranproteine, bilden einen der wichtigsten Grundsteine des Lebens. Schon im Jahre 1838 wurde ihre Bedeutung klar, als sie ihren Namen von J. Berzelius bekamen, welcher soviel bedeutet, wie das Erste oder das Wichtigste. Die Wichtigkeit der Proteine für einen Organismus macht vor allem ihre Funktion aus, welche durch die spezielle 3D-Struktur des Proteins bestimmt wird. Diese 3D-Struktur wird durch die wohldefinierte Abfolge von Aminosäuren bedingt und bildet somit ihre Grundlage. Für die Faltung dieser 3D-Struktur gibt es einen nativen Zustand, welcher den strukturell stabilsten und energetisch günstigsten Zustand widerspiegelt.

Es gibt viele Verfahren und Methoden, um Einblicke in das Verhalten von Proteinen zu erlangen, allerdings sind nicht alle Verfahren sowohl für globuläre als auch Membranproteine anwendbar. Die Anzahl der Methoden, welche einen direkten Erkenntnisgewinn aus dem Experiment mit dem Membranprotein ziehen, ist begrenzt. Zumeist besteht auch eine Divergenz zwischen den im Experiment erhaltenen und den beobachteten oder vorhergesagten Daten. Allerdings sind nicht nur die Methoden nicht unbegrenzt übertragbar, zusätzlich unterscheidet sich auch die Handhabbarkeit von Membranproteinen grundlegend von der globulärer Proteine.

Eines der Experimente, welche einen sofortigen Erkenntnisgewinn erzeugt, ist das sogenannten Single Molecule Force Spectroscopy (SMFS) Experiment. Dieses Experiment gibt Einblicke in die Stabilität von α -helikalen Membranproteinen und ist, bedingt durch die Eigenschaften von Membranproteinen, sehr kosten- und zeitintensiv. Auch die Auswertung und Übertragung der erhaltenen Daten auf das biologische System sind schwierig. Aus diesen Gründen befasst sich diese Arbeit mit der Erstellung eines Vorhersagealgorithmus, welcher das SMFS-Experiment simulieren und die Ergebnisse sammeln soll. Des Weiteren soll in dieser Arbeit versucht werden, die Lücke zwischen Experiment und Theorie zu schließen und eine Übertragung auf das biologische System möglich zu machen.

Dabei gliedert sich die Arbeit in folgende Kapitel:

Kapitel 1 fungiert als Einleitung und wird alle nötigen Grundlagen legen, um die Arbeit in ihrer Gänze verstehen und nachvollziehen zu können. Kapitel 2 wird den in dieser Arbeit zu Grunde liegenden und entwickelten Algorithmus beschreiben. In Kapitel 3 und 4 werden anschließend die erzeugten Ergebnisse gesammelt, zusammengefasst und evaluiert. Kapitel 5 wird sich der molekularbiologischen Interpretation widmen und einen Zusammenhang zwischen Theorie und Praxis herstellen. Im Abschluss werden

Kapitel 6, 7 und 8 eine Diskussion der Ergebnisse präsentieren, anschließend alles zusammenfassen und einen Ausblick geben.

1 Einleitung

Dieser einführende Abschnitt der Arbeit befasst sich mit den Grundlagen und Methoden, die benötigt werden, um diese Arbeit in ihrer Gänze verstehen und nachvollziehen zu können. Es werden unter anderem die grundlegenden Eigenschaften von Proteinen aufgezeigt, um anschließend die auf diese angewendeten Methoden zu erläutern. Dabei wird spezifischer auf Membranproteine eingegangen, da diese für die vorliegende Arbeit von Bedeutung waren.

1.1 Proteine: Aufbau und Funktion

Proteine sind Moleküle, welche sich aus einer spezifischen Abfolge von Aminosäuren zusammensetzen. Hierbei erfüllen sie sehr viele Aufgaben innerhalb der Zelle bzw. des Organismus, wie z. B. Energieproduktion oder auch den Stoffaustausch zwischen Zellen. Es wird angenommen, dass Proteine seit Beginn an der Entstehung des Lebens beteiligt sind. Eine Bestätigung dieser These zeigen die Experimente von Miller [\[MIL53\]](#), welcher nachwies, dass schon in der sogenannten „Ursuppe“ Aminosäuren und Proteine vorhanden waren und somit zu den ersten organischen Molekülen gehören.

Weiterhin gibt es zum Beispiel Tunnelproteine, welche Natriumionen unter ATP-Verbrauch aus der Zelle heraus und dafür Kaliumionen in die Zelle hinein pumpen. Bereits im Jahre 1957 wurde die Natrium-Kalium-ATPase von Jens Christian Skou entdeckt [\[SKO57\]](#). Auch die Signalweitergabe innerhalb der Zelle ist eine der wichtigen Rollen, welche von Proteinen übernommen wird. Weiterhin werden zum Beispiel der Geißelschlag bei der Fortbewegung von Organismen durch Proteinen bestimmt.

Bei der Betrachtung der Funktionen von Proteinen gibt es allerdings nicht nur welche, die selbst aktiv sind, sondern auch solche, die an andere Moleküle oder die DNS binden und diese somit markieren oder deren Konformation ändern, sodass andere Proteine ihrer Funktion nachkommen können. Ein Beispiel hierfür sind Antikörper, die an spezifische Virusproteine binden und diese somit für das Immunsystem markieren. Für die Konformationsänderung durch spezifische Proteine sind Chaperone zu erwähnen, welche neu synthetisierten Proteinen helfen, sich korrekt im Raum zu falten [\[Buc02\]](#).

Einen wichtigen Aspekt bei dem Aufbau von Proteinen bildet das sogenannte Rückgrat oder Backbone des Proteins. Dieses bezeichnet die kontinuierlich kovalent gebundenen Kohlenstoffatome des Proteins bzw. Makromoleküls. Diese Kohlenstoffatome werden auch C_{α} -Atome genannt.

In dieser Arbeit wird ein besonderes Augenmerk auf Membranproteine gelegt, da diese für diese Arbeit essentiell sind. Neben der bereits erwähnten Natrium-Kalium-ATPase gibt es noch viele weitere membrandurchspannende Proteine, welchen in dieser Arbeit, insbesondere α -helikalen, ein besonderes Interesse zukommt. α -helikale Membranproteine sind Proteine, welche mit ihren α -Helices die Membran durchspannen. Das Bei-

spiel der Natrium-Kalium-ATPase zeigt, dass Membranproteine vor allem die Aufgabe von Stoffaustausch und Stoffweitergabe haben. Im Gegensatz zu den globulären, wasserlöslichen Proteinstrukturen, von welchen bereits über 84000 bekannt sind [RBB⁺13], gibt es nur knapp 1700 Einträge in der Protein Data Bank of Transmembrane Proteins, der entsprechenden Datenbank für transmembrane Proteine [KST13], weshalb es umso wichtiger ist, Erkenntnisse über Membranproteine zu gewinnen.

1.2 3D-Struktur und Strukturaufklärung

Da diese Arbeit die energetischen und strukturellen Eigenschaften von α -helikalen Membranproteinen nutzt, ist es essentiell, die vollständige 3D-Struktur des betreffenden Proteins zu kennen. Das bedeutet, dass für jedes Atom einer Aminosäure dessen genaue Lage und Position im Raum bekannt sein muss. Um diese Daten zu generieren gibt es zwei gängige Methoden, welche im folgenden erläutert werden.

1.2.1 Röntgenstrukturanalyse

Dieses Verfahren wird auch Kristallstrukturanalyse [NH97a] genannt und bedient sich bestimmter Strahlung und eines Proteinkristalls. Bei der Strahlung handelt es sich entweder um monochromatische Röntgenstrahlung, Neutronenstrahlung oder Synchrotronstrahlung. Trifft diese Strahlung auf den Kristall des Proteins erhält man ein proteinspezifisches Beugungsmuster. Genau betrachtet wird hierbei nicht die Proteinstruktur bestimmt, sondern die Elektronenverteilung, da diese mit der Strahlung wechselwirkt. In die so erhaltene Elektronendichtekarte wird nun die Aminosäuresequenz des untersuchten Proteinkristalls eingepasst und verschoben, bis die Struktur die zuvor erhaltenen Elektronenverteilung plausibel erklärt.

1.2.2 Kernspinresonanzspektroskopie (NMR)

Das zweite Verfahren ist eine spektroskopische Methode, bei welcher die elektronische Umgebung einzelner Atome und deren Interaktionen mit Nachbaramen untersucht wird. Grundlage für die Möglichkeit der 3D-Strukturaufklärung von Proteinen bildet die magnetische Kernresonanz [NH97b]. Dies ist eine resonante Wechselwirkung des magnetischen Moments der Atomkerne innerhalb der Probe, welche auftritt, wenn man sie unter den Einfluss eines starken statischen Magnetfelds und einem hochfrequenten magnetischen Wechselfelds stellt. Hierbei gilt es unbedingt zu beachten, dass diese Methode nur für Isotope verwendbar ist, die im Grundzustand ein magnetisches Moment besitzen, d. h. deren Kernspin $\neq 0$ ist. Hier sind für Proteine als Beispiele zu nennen: ^1H , ^{13}C , ^{15}N , ^{17}O .

1.3 Theoretische und Praktische Methoden

Da sich ein Teil dieser Arbeit mit dem Entwickeln eines Algorithmus befasst, sind zunächst theoretische und praktische Methoden zu erläutern, die das Verständnis erleichtern sollen. Diese werden zum Beispiel benötigt, um das benötigte Dateiformat zu erzeugen und mit den richtigen Informationen zu versehen oder um das Protein künstlich zu verkürzen und von der so neu gewonnen Proteinstruktur die veränderten Energie- und Kontaktprofile zu erstellen. Zum Schluss folgt eine Erläuterung des Experiments, welches simuliert und vorhergesagt werden soll.

1.3.1 Anisotropic Network Model (ANM)

Mit Hilfe des ANM [DAB00] [ADJ⁺01] kann die Beziehung zwischen Funktion und Dynamik innerhalb eines Proteins aufgedeckt werden. In dem Modell werden biologische Makromoleküle, zu welchen auch Proteine gehören, als Masse-Feder-System ausgehend von C_α -Atomen, dargestellt. Die C_α -Atome werden hierbei von Knoten repräsentiert, während die Interaktionen zwischen diesen durch die Federn symbolisiert werden. Das Gesamtpotential des Systems ist gleich der Summe der harmonischen Potentiale zwischen den interagierenden C_α -Atomen. Es gibt hierbei nur einen Freiheitsgrad, der die innere Bewegung der Feder zwischen zwei Knoten beschreibt. Dieser korrespondiert mit der Ausdehnung und Stauchung der durch die Feder verbundenen Knoten. Die Anwendung dieses Modells auf die Proteinstruktur und auf die künstlich verkürzten Proteinstrukturen (vgl. 1.3.5) kann eine Abbildung der entstehenden Bewegungen innerhalb des Proteins sein, die entstehen, wenn das Protein während des SMFS-Experiments aus der Membran herausgelöst wird. Die folgende Abbildung 1.1 zeigt ein Beispiel für eine solche Bewegungspfad (Trajektorie).

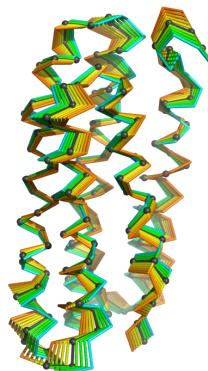


Abbildung 1.1: Beispiel einer durch ANM berechneten Trajektorie, welche die mögliche Bewegung der Proteinstruktur abbildet. In grau ist hierbei die Backbone der Proteinstruktur zu erkennen. Die unterschiedlich farbigen Strukturen zeigen jeweils eine Konformation zu unterschiedlichen Zeitpunkten [HAD13].

1.3.2 Energieprofil und Kontaktprofil

Das Energieprofil ist proteinspezifisch und bedient sich der spezifischen Lösungsenergien der Aminosäuren des Proteins. Gleichung 1.1 beschreibt den Energiewert, und Gleichung 1.2 einen aminosäurespezifischen Wert, welcher sich aus dem Vorkommen der Aminosäure innerhalb bzw. außerhalb der Membran zusammensetzt.

$$(1.1) \quad e_i = -\ln \left(\frac{f_{bur,i}}{f_{exp,i}} \right) + k_i$$

$$(1.2) \quad k_i = -\ln \left(\frac{f_{i,inner-membrane}}{f_{i,outer-membrane}} \right)$$

Um die betrachtete Aminosäure wird nun eine 8Å-Sphäre gelegt, um die Interaktionen mit anderen Aminosäureresten zu bestimmen. Gleichung 1.3 zeigt dabei, ob zwischen zwei Aminosäureresten i und j eine Interaktion besteht. Besteht eine Interaktion, ist der Wert 1, sonst 0.

$$(1.3) \quad g(i, j) = \begin{cases} 1 & \forall \| C_{\beta,i} - C_{\beta,j} \| \leq 8\text{\AA} \\ 0 & \text{else} \end{cases}$$

Für eine bestehende Interaktion ergibt sich somit folgender Energiewert für die Aminosäure i.

$$(1.4) \quad e_{i,j} = e_i + e_j$$

In Abhängigkeit von dieser Umgebung, welche durch die 8Å-Sphäre symbolisiert wird, wird der spezifische Energiewert der Aminosäure, unter Einbeziehung der in dieser Sphäre liegenden C_{β} -Atome, für Glycin C_{α} -Atome, und deren spezifische Energie, als Summe der Lösungsenergien dargestellt.

$$(1.5) \quad E_i = \sum_{j \in \text{Protein} | j \neq i} g(i, j) [e_{i,j}]$$

Das Kontaktprofil verwendet die 8Å-Sphäre und summiert die Anzahl der auftretenden C_{β} -Atome, für Glycin C_{α} -Atome, innerhalb der 8Å-Sphäre, um den für jede Aminosäure des Proteins spezifischen Wert zu bestimmen. Dabei verwendet das Kontaktprofil die

Gleichung 1.3, um eine Interaktion zwischen zwei Aminosäureresten zu bestimmen. Die Gleichung 1.6 bildet, analog zur Gleichung 1.5 beim Energieprofil, die Summe über alle Interaktionen, um den spezifischen Kontaktwert für jede Aminosäure zu berechnen.

$$(1.6) \quad C_i = \sum_{j \in \text{Protein} | j \neq i} g(i, j)$$

1.3.3 3D-Strukturvorhersage

In Kapitel 1.2 wurde die Wichtigkeit von aufgeklärten 3D-Strukturen von Proteinen bereits eingehend erläutert. Da es sehr schwierig und auch kosten- bzw. zeitintensiv ist, ein Protein in seiner 3D-Struktur aufzuklären, gibt es von vielen Proteinen keine aufgeklärte 3D-Struktur. Zu diesen nicht aufgeklärten Proteinen zählen auch sehr viele Mutanten von Proteinen. Da die Betrachtung von Mutanten oder nicht aufgeklärten Proteinen allerdings durchaus benötigt wird, gibt es die Möglichkeit diese 3D-Strukturen modellieren zu lassen. Die Modellierung die in dieser Arbeit zum Einsatz kommt ist die sogenannte Homologiemodellierung. Dieses Verfahren bedient sich einer Vorlage (Template), nach dessen Vorlage das Unbekannte Protein modelliert wird. Ein Teil der nicht aufgeklärten Proteine in dieser Arbeit, wurden durch dieses Verfahren der SwissProt [ABKS06] [KAK⁺09] modelliert. Der andere Teil wurde aus der ModBase Datenbank entnommen, welche durch die Homologiemodellierung mit Hilfe von Modeller [EWMR⁺06] erzeugt wurden.

1.3.4 Single Molecule Force Spectroscopy (SMFS)

Das SMFS-Experiment ist ein Verfahren, um Energiebarrieren innerhalb von transmembranen Proteinen bestimmen zu können. In diesem Verfahren werden die Proteine in Lösung auf die sogenannte 'native purple membrane' aufgetragen [OS74] und von 'freshly cleaved mica' [MAE97] adsorbiert. Anschließend wird ein einzelnes Molekül zum Kantilever geführt. Dieses geht anschließend eine Wechselwirkung mit der Spitze des Kantilevers ein. Auf das so adhäsierende Molekül wird nun eine externe Kraft ausgeübt, um es kontrolliert zu entfalten. Die Abbildung 1.2 stellt den Aufbau des Experimentes dar.

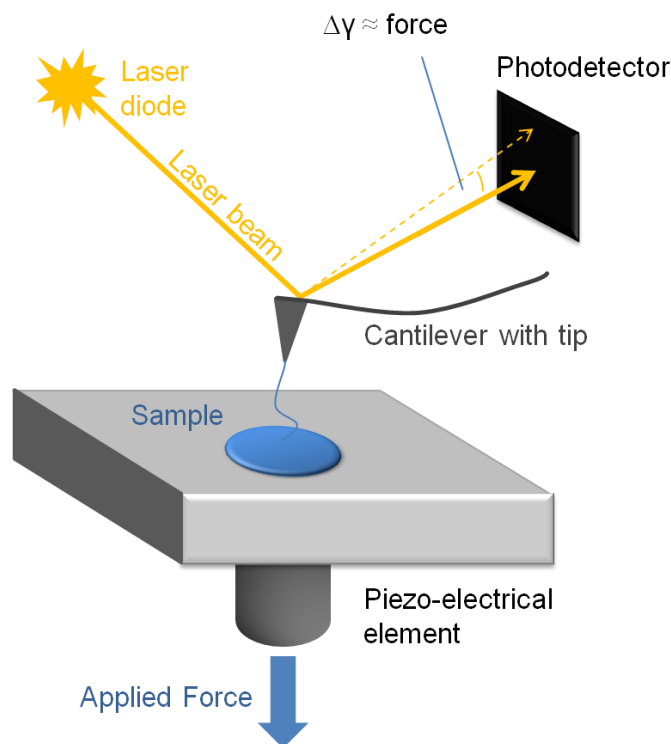


Abbildung 1.2: Beispielhafter Aufbau eines SMFS-Versuchsvorgangs. Zu sehen sind das elektrische Piezoelement, auf dessen Oberfläche die Probe aufgebracht wird. Des Weiteren sieht man den Kantilever mit seiner Spitze. Die Laserdiode und der Photodetektor sind nötig, um den Abstand bei eingebrachter Kraft akkurat zu messen [HAD13].

Mit Hilfe dieses Verfahrens generiert man für jedes Molekül eine Kraft-Abstands-Kurve (engl.: Force-Distance-Curve 'F-D'). Da man dieses Verfahren im Hochdurchsatz durchführt, werden mehrere Kurven erzeugt und übereinander gelagert. Das Ergebnis zeigt Abbildung 1.3.

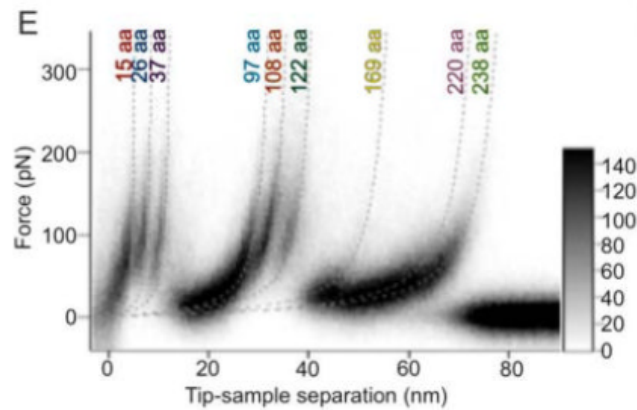


Abbildung 1.3: Beispielhafte Kraft-Abstands-Kurve für das C-Terminal gezogene Bacteriorhodopsin mit eingetragener Korrelation zur Aminosäure, berechnet durch das 'Wormlike-chain-model' [KCMP10].

Diese so erhaltenen Ergebnisse werden als SMFS-Peak bezeichnet und können nun durch das 'Wormlike-chain-model' auf die betreffenden Aminosäuren abgebildet werden. Diese so ermittelten Aminosäuren werden als Entfaltungsbarriere bezeichnet.

Zum Verständnis dient Abbildung 1.4, welche einen idealisierten Entfaltungsprozess eines Proteins beim SMFS-Vorgang zeigt.

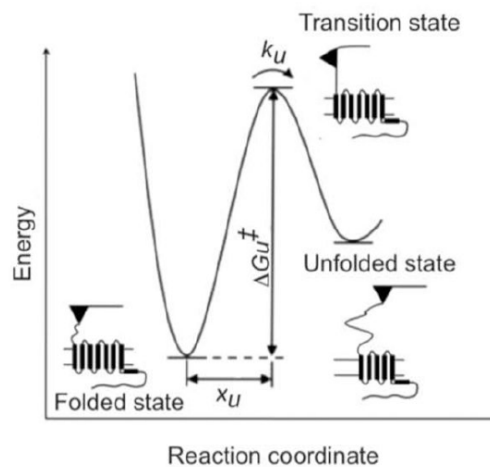


Abbildung 1.4: Idealisiertes Entfaltungsmodell für ein stabiles Strukturelement. Der Parameter x_U bezeichnet den Abstand zwischen dem Energieminimum des gefalteten Zustands und des Übergangszustands. Der Parameter k_U bezeichnet die Rate des Entfaltens vom gefalteten Segment unter Abwesenheit von Kräften. ΔG_U^\ddagger bezeichnet die Höhe der Barriere des Übergangszustands [KCMP10].

Die folgende Abbildung 1.5 stellt eine Zusammenfassung aller bereits erwähnten Schritten des Experiments dar. Es werden sowohl die Daten des SMFS-Experiments und somit die SMFS-Peaks gezeigt als auch die Zuordnung zu den entsprechenden Aminosäuren durch das Wormlike-chain-model. Auch wird aufgezeigt, dass es verschiedene Entfaltungswege bei dem Vorgang des kontrollierten Entfaltens gibt.

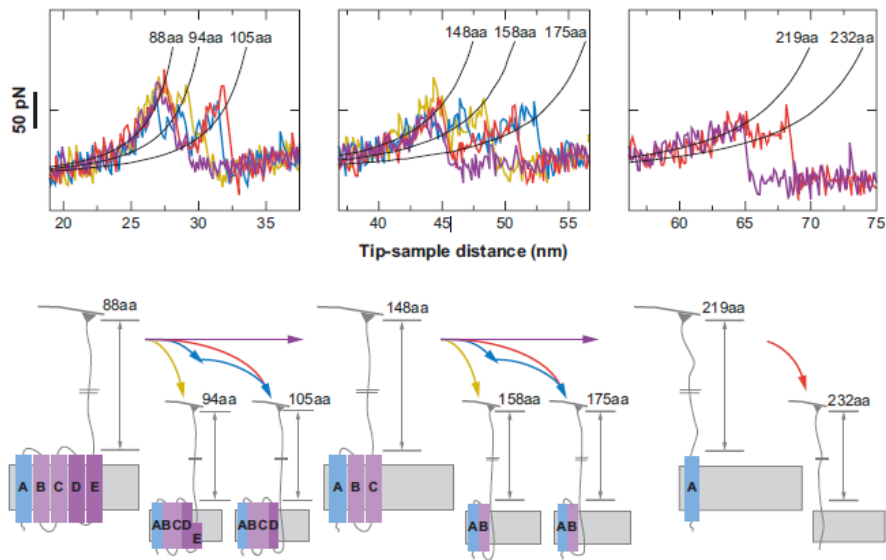


Abbildung 1.5: Zusammenfassende Darstellung des SMFS-Experiments. Der obere Teil zeigt die Daten des Experiments mit den entsprechenden SMFS-Peaks und den dazugehörigen Aminosäureresten, welche durch das Wormlike-chain-Modell berechnet wurden. Der untere Teil stellt schematisch die verschiedenen Entfaltungswege mit den entsprechenden Entfaltungsbarrieren, schematisch aufgetragen gegen die Helices des Proteins, dar [KJSM07].

1.3.5 In silico-Simulation des SMFS-Entfaltens

Um das in 1.3.4 erläuterte Verfahren in einem einfachen, computergestützten Modell abbilden zu können, bedient sich diese Arbeit einer einfachen Simulation des Experiments. Hierbei wird die spezifische Datei der Protein Data Bank (PDB) des Proteins, welche die Lage aller Atome im Raum enthält, in einzelnen Aminosäureschritten künstlich verkürzt, um das kontrollierte Entfalten aus der Membran zu simulieren. Das bedeutet, dass das Protein am Computer schrittweise um eine Aminosäure verkürzt wird, beginnend bei einem der beiden Termini, um eine neue, veränderte PDB-Datei zu erzeugen.

1.4 Verwendete Proteine

Eine weitere Grundlage dieser Arbeit bilden aufwändig aufgeklärte α -helikale Membranproteine. Wie in Kapitel 1.1 bereits erwähnt, ist der Datenbankbestand von Membranproteinen im Vergleich zu globulären Proteinen sehr gering. Dies liegt unter anderem an dem aufwändigen Verfahren, um eine 3D-Struktur eines Membranproteins zu erhalten. Alle Abbildungen von Proteinen in dieser Arbeit wurden mit Hilfe der Software PyMOL [Sch10] erstellt.

1.4.1 Bacteriorhodopsin

Das Protein Bacteriorhodopsin, mit seiner spezifischen Protein Data Bank Identifizierung (PDB-ID) 1BRR [ESLO98], ist ein integrales Membranprotein mit sieben transmembranen Helices und dient als Protonenpumpe.

Laut dem Eintrag P02945 der UniprotKB besitzt das Protein Bacteriorhodopsin eine Länge von 262 Aminosäuren, beginnend bei der ersten. Hierbei sind die Reste 1 – 13 ein Propeptid, was bedeutet, dass das eigentliche Proteine aus den Aminosäuren 14 bis 262 besteht. Das in der Arbeit verwendete Protein mit der PDB-ID 1BRR besteht allerdings nur aus den Resten 3 bis 232, von denen nur die Reste 3 bis 230 in die Berechnungen eingehen. Die nachfolgende Abbildung zeigt das Protein in seiner 3D-Struktur.

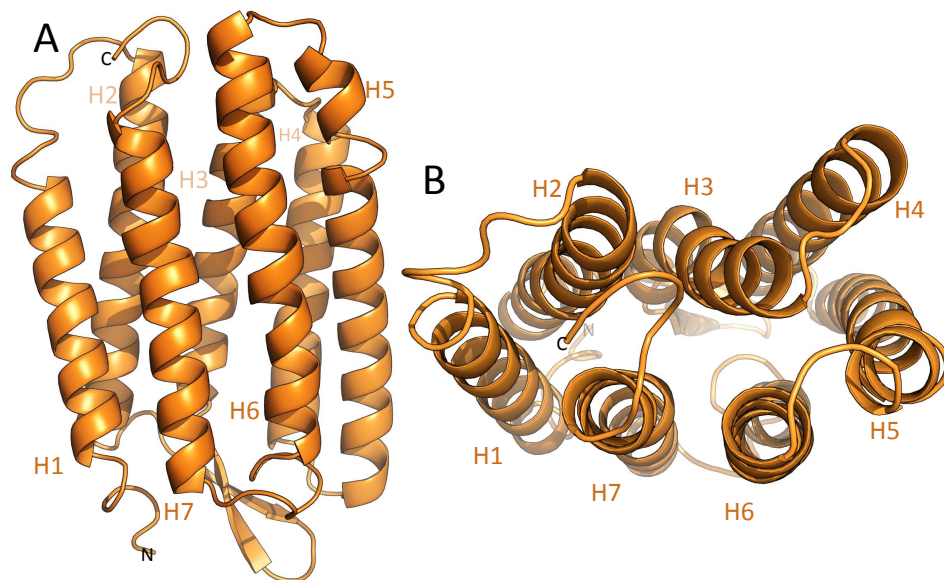


Abbildung 1.6: Cartoondarstellung der Kette A des Proteins Bacteriorhodopsin mit der PDB-ID 1BRR. **A** zeigt die Seiten- und **B** die Draufsicht des Proteins. Mit N und C sind die jeweiligen Termini des Proteins bezeichnet. Die einzelnen Helices sind mit H bezeichnet und ausgehend vom N-Terminus durchnummeriert.

1.4.2 Halorhodopsin

Halorhodopsin mit der PDB-ID 1E12 [[KBEO00](#)] ist eine, für Chloridionen spezifische, lichtinduzierte Ionenpumpe und besteht aus sieben transmembranen Helices und gehört somit zur Gruppe der integralen Membranproteinen.

Laut dem Eintrag P16102 der UniprotKB besitzt das Protein Halorhodopsin eine Länge von 274 Aminosäuren, beginnend bei der ersten. Hierbei sind die Reste 1 – 21 ein Propeptid, was bedeutet, dass das eigentliche Proteine aus den Aminosäuren 22 bis 274 besteht. Das in der Arbeit verwendete Protein mit der PDB-ID 1E12 besteht allerdings nur aus den Resten 23 bis 262 von denen nur die Reste 23 bis 260 in die Berechnungen eingehen. Die nachfolgende Abbildung zeigt das Protein in seiner 3D-Struktur.

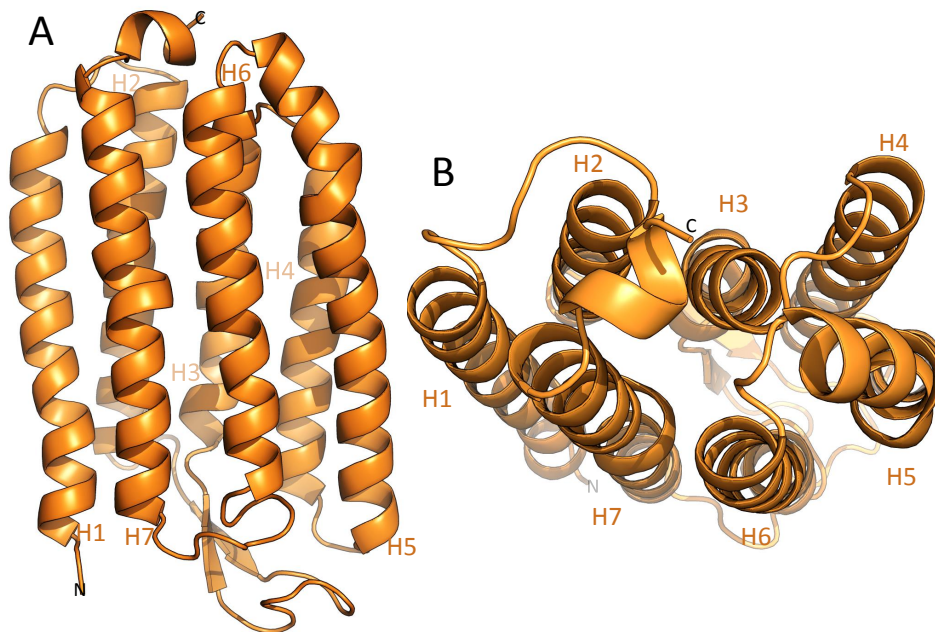


Abbildung 1.7: Cartoondarstellung der Kette A des Proteins Halorhodopsin mit der PDB-ID 1E12. **A** zeigt die Seiten- und **B** die Draufsicht des Proteins. Mit N und C sind die jeweiligen Termini des Proteins bezeichnet. Die einzelnen Helices sind mit H bezeichnet und ausgehend vom N-Terminus durchnummeriert.

1.4.3 Humanes Aquaporin 1

Das als Wassertunnel dienende humane Aquaporin 1 mit der PDB-ID 1FQY [MMH⁺00] ist ein integrales Membranprotein mit sieben transmembranen Helices.

Laut dem Eintrag P29972 der UniprotKB besitzt das Protein humane Aquaporin 1 eine Länge von 269 Aminosäuren, beginnend bei 1. Hierbei wird das initialisierende Methionin entfernt, was bedeutet, dass das eigentliche Proteine aus den Aminosäuren 2 bis 269 besteht. Das in der Arbeit verwendete Protein mit der PDB-ID 1FQY besteht allerdings nur aus den Resten 8 bis 233 von denen nur die Reste 8 bis 231 in die Berechnungen eingehen. Für dieses Proteine musste bei den experimentellen Daten die ermittelte Entfaltungsbarriere bei der Aminosäure 235 aus der Betrachtung herausgenommen werden, da das hier verwendete Protein im Vergleich zur UniProt kürzer und damit die besagte Barriere nicht vorhanden ist. Dies führt dazu, dass eine Vorhersage dieser Barriere nicht möglich ist, weshalb sie aus der Betrachtung herausgenommen wurde. Die nachfolgende Abbildung zeigt das Protein in seiner 3D-Struktur.

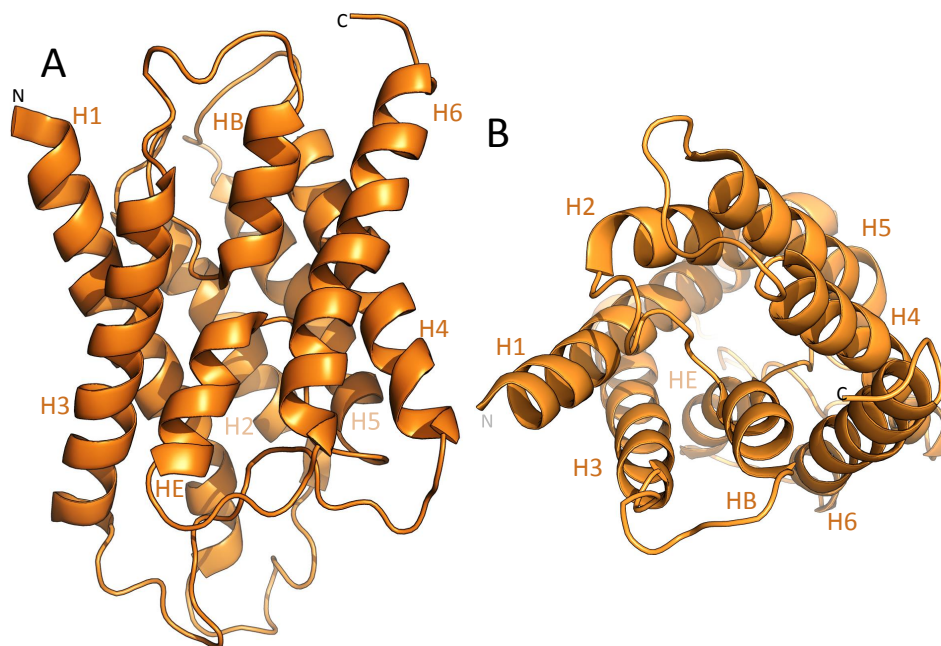


Abbildung 1.8: Cartoondarstellung der Kette A des Proteins Aquaporin mit der PDB-ID 1FQY. **A** zeigt die Seiten- und **B** die Draufsicht des Proteins. Mit N und C sind die jeweiligen Termini des Proteins bezeichnet. Die einzelnen Helices sind mit H bezeichnet und ausgehend vom N-Terminus durchnummeriert. Die Helices HE und HB bilden hierbei eine proteinspezifische Besonderheit.

1.4.4 Natives Rinderrhodopsin

Das Rinderrhodopsin mit der PDB-ID 1U19 [OSB⁺04] ist die Rhodopsinvariante des Rindes und dient, ebenso wie bei dem Menschen und anderen Wirbeltieren, als Sehpigment der Netzhaut. Es handelt sich hierbei um den ersten G-Protein gekoppelten Rezeptor (GCPR), dessen Struktur mittels Röntgenstrukturanalyse aufgeklärt wurde. Laut dem Eintrag P02699 der UniprotKB besitzt das Protein Halorhodopsin eine Länge von 348 Aminosäuren, beginnend bei der ersten. Das in der Arbeit verwendete Protein mit der PDB-ID 1U19 besteht aus allen Aminosäuren von denen die Reste 1 bis 346 in die Berechnungen eingehen. Die nachfolgende Abbildung zeigt das Protein in seiner 3D-Struktur.

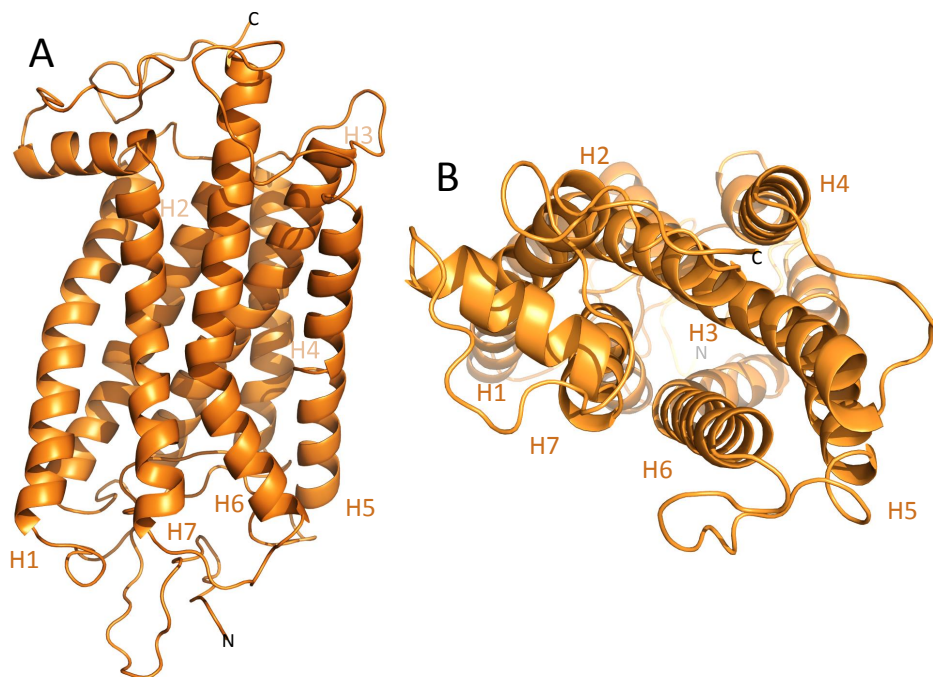


Abbildung 1.9: Cartoondarstellung der Kette A des Proteins Rinderrhodopsin mit der PDB-ID 1U19. **A** zeigt die Seiten- und **B** die Draufsicht des Proteins. Mit N und C sind die jeweiligen Termini des Proteins bezeichnet. Die einzelnen Helices sind mit H bezeichnet und ausgehend vom N-Terminus durchnummeriert.

1.4.5 Mutationen

Im Rahmen der vorliegenden Arbeit werden einige Mutationen der Proteine Bacterio- und Halorhodopsin betrachtet. Bei diesen handelt es sich ausschließlich um sogenannte Punktmutationen, das heißt eine spezielle Aminosäure wurde durch eine andere ausgetauscht bzw. ersetzt. Für die Benennung dieser Mutationen gibt es eine einheitliche Struktur. Diese allgemein verwendete Darstellung sieht wie folgt aus: B100X. Hinter dieser allgemeinen Darstellung verbirgt sich die Aussage, dass die zuerst genannte Aminosäure, im Beispiel Aminosäure B, an der spezifischen Stelle der Aminosäuresequenz, im Beispiel Position 100, durch die nach der Zahl genannten Aminosäure ersetzt wurde, im Beispiel Aminosäure X. Zusammengefasst bedeutet das für den Mutanten, dass die Aminosäure B an der Position 100 durch die Aminosäure X ersetzt wurde.

1.5 Gütemaße

Im folgenden werden die in dieser Arbeit verwendeten Gütemaße vorgestellt. Hierbei werden sowohl eine Klassifizierungsmöglichkeit als auch zwei verschiedene Gütemaße erläutert. Diese zwei verschiedenen Gütemaße werden verwendet, um insgesamt eine bessere Aussage über die Güte des entwickelten Algorithmus treffen zu können, da beide unterschiedlich von den Ergebnissen beeinflusst werden.

1.5.1 Konfusionsmatrix

Die Konfusionsmatrix ist eine kleine 2×2 -Matrix und visualisiert die Klassenzuweisung eines binären Klassifikators in Abhängigkeit gegebener Evaluierungsdaten und deren bekannten Klassen. Hierbei ergeben sich vier Klassifizierungsfälle: korrekt klassifizierte positive Klassen (TP), falsch klassifizierte positive Klassen (FN), richtig klassifizierte negative Klassen (TN) und falsch klassifizierte negative Klassen (FP). Eine intuitive Darstellung der Klassifikationsergebnisse kann mit Hilfe der Konfusionsmatrix erfolgen (siehe Tabelle 1.1).

Tabelle 1.1: Allgemeine Konfusionsmatrix [KP98]

		Experiment	
		positiv	negativ
Vorhersage	positiv	TP	FP
	negativ	FN	TN

Mit Hilfe dieser Matrix lässt sich schnell und einfach ein Überblick über die Güte einer Vorhersage gewinnen, da die Summe aus TP und FN gleich der insgesamt vorherzusagenden Entfaltungsbarrieren ist.

1.5.2 Matthews Correlation Coefficient (MCC)

Der Matthews Correlation Coefficient (MCC) [Pow11] kann, unter Verwendung der TP, der FP, der TN und der FN eine Aussage über die Güte des Vorhersagealgorithmus treffen. Er wird wie in Gleichung 1.7 berechnet:

$$(1.7) \quad MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

Der MCC kann Werte zwischen -1 und 1 annehmen, wobei -1 eine umgekehrte Korrelation zwischen Experiment und Vorhersage, 0 eine zufällige Beobachtung und 1 die perfekte Vorhersage beschreibt.

1.5.3 F1-Maß (F1)

Das F1-Maß (F1) [Pow11] verwendet die Beziehung zwischen Genauigkeit und Trefferquote, um eine Aussage über die Güte zu treffen. Die einzelnen Komponenten werden wie folgt berechnet:

$$(1.8) \quad \text{Genauigkeit} = \frac{TP}{(TP + FP)}$$

$$(1.9) \quad \text{Trefferquote} = \frac{TP}{(TP + FN)}$$

$$(1.10) \quad F1 = 2 \times \frac{(\text{Genauigkeit} \times \text{Trefferquote})}{(\text{Genauigkeit} + \text{Trefferquote})}$$

Das F1-Maß kann dabei nur Werte zwischen 0 und 1 annehmen und kann als gewichteter Durchschnitt von Genauigkeit und Trefferquote angesehen werden. Hierbei steht 1 für den besten und 0 für den schlechtesten Wert.

2 Barrieren-Vorhersagealgorithmus

In diesem Abschnitt der Arbeit wird der entwickelte und zur Barrierenvorhersage verwendete Algorithmus vorgestellt und erläutert. Des Weiteren wird der Algorithmus nicht nur in Text sondern auch als Fließschema dargestellt und erläutert, um eine bessere Reproduzierbarkeit des Algorithmus und somit der hier vorgestellten Ergebnisse zu erhalten.

2.1 Pseudocode

Nachdem jedes Protein künstlich verkürzt wurde, um das kontrollierte Entfalten aus der Membran zu simulieren, wird für jede so erhaltenen Proteinstruktur (State) mittels ANM, eine diskrete Bewegungstrajektorie (Mode) erstellt, um das Experiment besser abbilden zu können. Anschließend wird für jede so erzeugte Proteinstruktur in der Bewegungstrajektorie das Energie- und Kontakprofil berechnet.

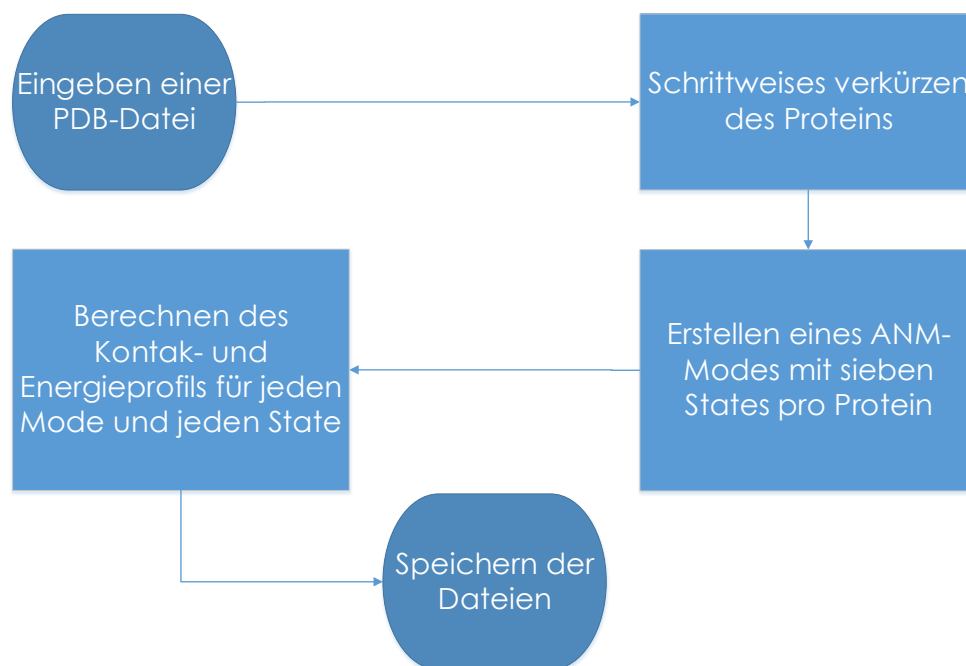


Abbildung 2.1: Darstellung der Funktionsweise des ersten Schritts des verwendeten Algorithmus als Fließdiagramm

Diese Kontakt- und Energieprofile werden einzeln, nacheinander eingelesen. Hierbei

wird für jede Datei und jedes Profil eine Normalisierung durchgeführt, was bedeutet, dass dem höchsten Wert eine 1 und dem niedrigsten Wert eine 0 zugeordnet wird. Die Werte für jede Datei werden für jede einzelne Aminosäure aufsummiert, worauf anschließend der Mittelwert der Aminosäure berechnet wird. Am Ende dieses Schrittes erhält man zwei Listen mit allen Aminosäuren des Proteins, wobei jeder Aminosäure in jeder der zwei Listen ein Wert zwischen 0 und 1 zugeordnet ist.

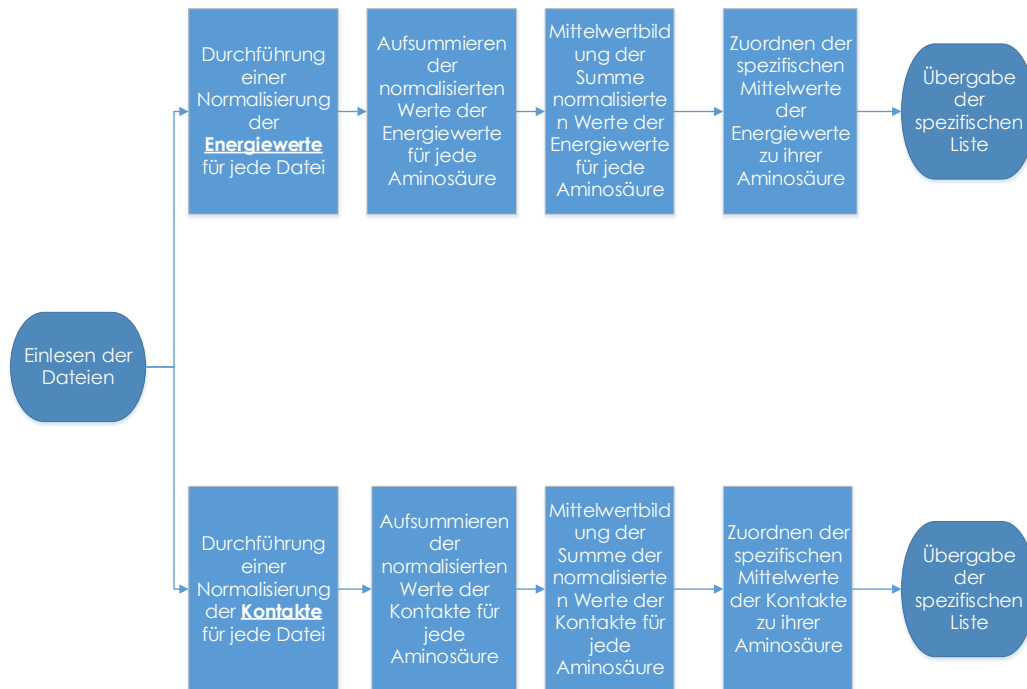


Abbildung 2.2: Darstellung der Funktionsweise des zweiten Schritts des verwendeten Algorithmus als Fließdiagramm

Der nächste Schritt (vgl. Abbildung 2.3) der Erstellung der Barrierenvorhersage ist nun das aminosäureweise Durchlaufen beider Listen. Hierbei wird überprüft, ob die spezifisch zugeordneten Werte der Aminosäure einen festgelegten, kombinierten Schwellwert unterschreiten und ob mindestens einer der Werte aus Kontakt- oder Energiewertliste, einen weiteren festgelegten Schwellwert unterschreitet (vgl. Abschnitt 2.2). Anhand des zweiten Schwellwertes kann festgestellt werden, ob es sich um eine strukturell oder energetisch vermittelte Barriere handelt. Dies ist notwendig, um Barrieren, welche außerhalb der Membran in Loops gelegen sind, vorhersagen zu können, aber auch solche, die in den Helices innerhalb der Membran gelegen sind.

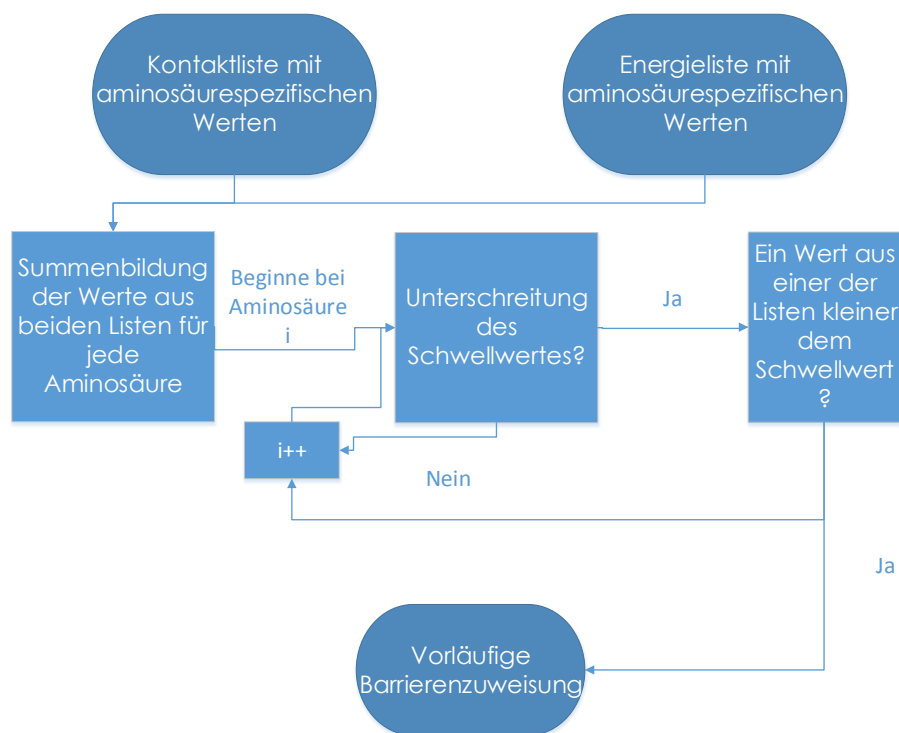


Abbildung 2.3: Darstellung der Funktionsweise des dritten Schritts des verwendeten Algorithmus als Fließdiagramm

Die Vorhersage wird nun in eine Binärsequenz überführt, d. h. eine Aminosäure, welche den Schwellwert unterschreitet, wird als 1 und eine, die den Schwellwert nicht unterschreitet als 0 dargestellt. Zur Verringerung des Rauschens innerhalb der Vorhersage, wird ein Fenster über die Binärsequenz geschoben, um in einem wohldefinierten Bereich Vorhersagen zusammenzufassen. Das bedeutet in diesem Falle, dass aus allen positiven Vorhersagen im Intervall der Mittelwert gebildet wird. Ist dieser Mittelwert nicht ganzzahlig, so wird er abgerundet. Dieses Fenster wird immer zur nächsten Vorhersage, entsprechend einer 1 in der Binärsequenz, weitergeschoben. Dieses Verfahren wird zweimal durchgeführt. Abschließend wird die letzte zusammengefasste Binärsequenz nach Positionen durchsucht, welche 1 sind, um die letztendliche Barrierenvorhersage auszugeben (vgl. Abbildung 2.4).

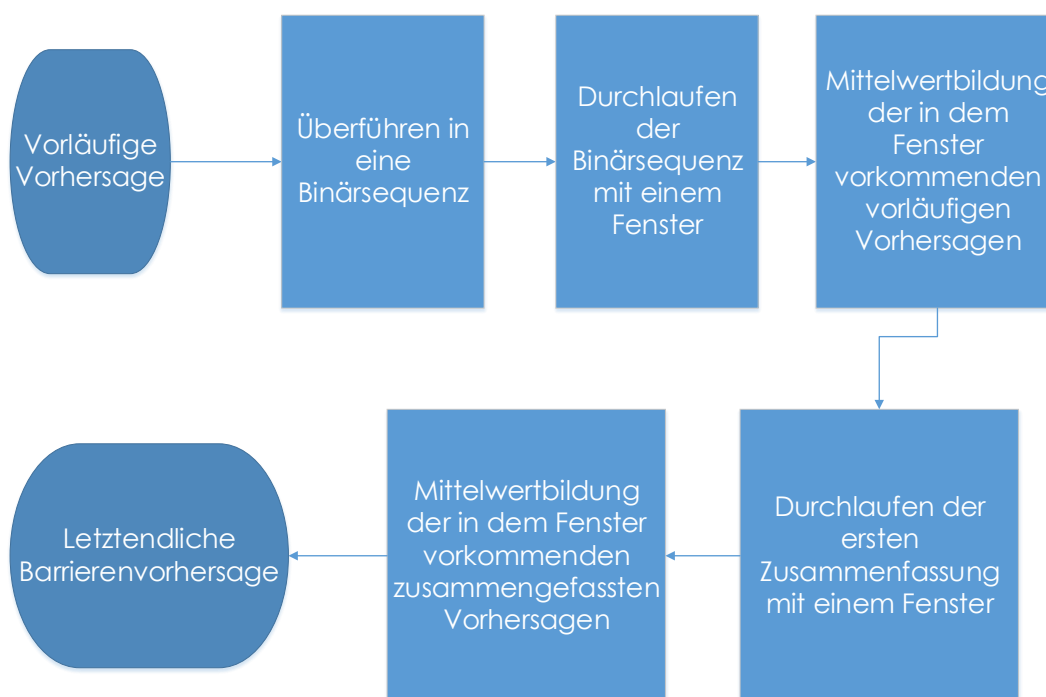


Abbildung 2.4: Darstellung der Funktionsweise des vierten Schrittes des verwendeten Algorithmus als Fließdiagramm.

Da aufgrund der vorhandenen Daten der Proteine die Coilregionen des C- bzw. N-Terminus immer vorhergesagt werden, korrigiert man diese Vorhersagen noch. Korrektur bedeutet in diesem Falle, dass die Vorhersage dieser Entfaltungsbarriere entfernt wird.

Um anschließend zu überprüfen, ob eine Vorhersage im Einklang mit den Experimenten steht, gibt es noch ein weiteres Fenster, mit dem überprüft wird, ob eine vorhergesagte Barriere auch im Experiment nachgewiesen wurde. Dieses Fenster wird auf die vorhergesagte Aminosäure gelegt und überprüft, ob sich eine oder mehrere experimentell ermittelte Barrieren innerhalb des Fensters befinden oder nicht.

2.2 Parametrisierung

Der verwendete Algorithmus nutzt insgesamt vier Parameter, welche im folgenden erläutert werden.

Der erste verwendete Parameter ist ein Schwellwert. Dieser setzt sich aus den normierten Werten aus der Energiewert- und Kontaktliste zusammen. Er kann zwischen 0 und 2 liegen (vgl. Kapitel 2.1). Damit eine Aminosäure als Barriere zunächst in Frage kommt, muss die Summe ihrer normierten Werte kleiner gleich 0,64 sein. Um zu gewährleisten, dass keine einmaligen Ausreißer mit einbezogen werden, gibt es danach noch einen zweiten Schwellwert. Dieser zweite Schwellwert bezieht sich nur noch auf

einen der beiden normalisierten Werte, kann somit nur zwischen 0 und 1 liegen und muss mindestens von einem der beiden Werte unterschritten werden. Dieser zweite Schwellwert wurde zusammen mit dem kombinierten ersten Schwellwert empirisch ermittelt und muss $\leq 0,31$ sein.

Die andern beiden Parameter sind Fenstergrößen. Das erste Fenster wird verwendet, um das Rauschen bei der Barrierenvorhersage zu vermindern, und wurde empirisch ermittelt. Dieses Fenster hat in Kombination mit den ersten beiden Parametern einen idealen Wert von 4. Das bedeutet, dass die betrachtete Aminosäure und die vier darauffolgenden zusammengefasst werden. Insgesamt werden damit fünf Aminosäuren zusammengefasst. Der letzte Parameter bezieht sich auf das Fenster, in welchem vorhergesagte Barrieren, bzw. deren Aminosäuren im Vergleich mit dem Experiment noch als richtige Vorhersage gelten. Das hierfür verwendete Fenster wurde auf ± 7 Aminosäuren festgelegt.

3 Ergebnisse

Dieser Abschnitt der Arbeit wird die Ergebnisse darlegen, welche mit Hilfe des in Kapitel 2 vorgestellten Algorithmus erstellt wurden. Da zu zwei der vier analysierten Proteine auch Mutanten betrachtet wurden, werden zunächst die Ergebnisse der Vorhersagen des Wildtyps des Proteins tabellarisch dargestellt, da diese zur Betrachtung der Güte in Kapitel 4 herangezogen werden. Die Vorhersagen der einzelnen betrachteten Mutationen werden immer im Vergleich zum intakten Protein dargestellt. Da die Ergebnisse des nativen Zustands, bezogen auf die Gütemaße MCC und F1, bedeutend besser waren als jene, die versuchen das Experiment zu simulieren, wurde für den Rest der Arbeit lediglich der native Zustand der Proteine verwendet (vgl. hierzu Kapitel 4). Das bedeutet in diesem Fall, dass die Simulation des SMFS-Experiments wie in Abbildung 2.1 beschrieben, nicht stattfindet, sondern lediglich für den Grundzustand des Proteins ein Energie- und Kontaktpprofil berechnet wird. Der restliche Ablauf des Algorithmus wird hiervon allerdings nicht beeinflusst und erfolgt genau so wie in den Abbildungen 2.2 bis 2.4 beschrieben.

3.1 Bacteriorhodopsin

Die Ergebnisse der Barrierenvorhersage werden direkt mit im Experiment ermittelten Barrieren verglichen [MKO⁺02], wobei das Fenster zur Bestimmung von richtig vorhergesagten Entfaltungsbarrieren beachtet werden muss (vgl. Kapitel 2.2).

Tabelle 3.1: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein. In der rechten Spalte sind die experimentellen Daten aufgetragen. Ein ' - ' in einer Spalte bedeutet, dass es für diesen Wert keine Vorhersage im Bereich des Fensters (vgl. Kapitel 2.2) oder keine experimentellen Daten gibt. Ein ' / ' in einer Spalte bedeutet, dass mehrere Daten miteinander vereinbar sind.

Vorhersage Bacteriorhodopsin	experimentelle Daten
10/17	15
24/30	27
48	-
60	-
72	72
88	89
105	99
116	-
140	142
154	153
154	159
185	-
197	199
210	211
226	221

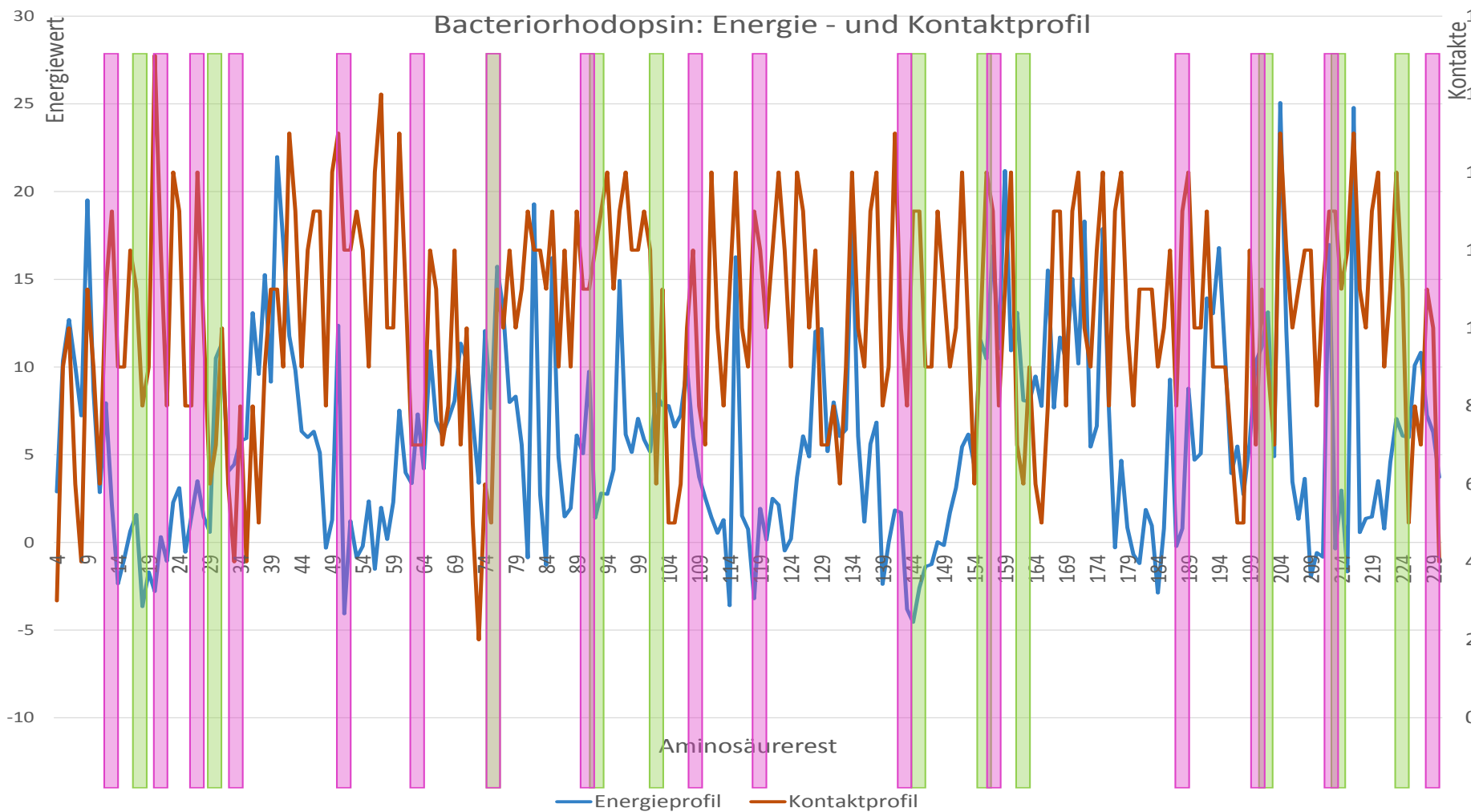


Abbildung 3.1: Darstellung der Ergebnisse der Barrierenvorhersage des Proteins Bacteriorhodopsin. In blau ist der Verlauf der Energiewerte zum jeweiligen Aminosäurerest gegen die linke y-Achse aufgetragen. In braun ist Verlauf der Kontaktanzahl zum jeweiligen Aminosäurerest gegen die rechte y-Achse aufgetragen. Die grünen Balken zeigen die experimentell ermittelten Barrieren, während die magentafarbenen Balken die vorhergesagten Aminosäurereste anzeigen. Grüne Balken mit magentafarbenem Rand zeigen eine genaue Übereinstimmung von Vorhersage und Experiment.

3.1.1 Mutationen

Für das Protein Bacteriorhodopsin wurden insgesamt 17 verschiedene Mutationen betrachtet, um den Einfluss dieser auf die Vorhersagen, welche mit Hilfe des in Kapitel 2 vorgestellten Algorithmus erstellt werden, zu untersuchen. Diese Betrachtung ist notwendig, um zum einen den Einflussbereich der Mutation überprüfen zu können und zum anderen die Möglichkeit zu überprüfen, ob es möglich ist, mit nur einer Mutation eine Barrierenvorhersage zu verursachen bzw. eine Vorhersage ausfallen zu lassen. Hierzu wurden sowohl aufgeklärte als auch modellierte Proteine betrachtet.

Um die Daten optimal vergleichen zu können, werden die Vorhersagen der Mutanten mit dem des nativen Proteins verglichen.

3.1.2 Mutante - D96N

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1C8R [LSR⁺99] zeichnet sich durch eine D96N Mutation aus. Die Aminosäure Asparaginsäure an der Position 96 wurde dabei durch die Aminosäure Asparagin ausgetauscht. Diese Mutation führt zu einer Verschiebung von Seitenketten in der Nähe des Retinals und einiger Wasserstoffbrückenbindungen, welche durch eine Veränderung der Säurekonstante entstehen.

Tabelle 3.2: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Ein ' - ' in einer Spalte bedeutet, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	27
30	30
48	-
60	60
72	73
88	86
105	110
116	-
140	138/144
154	-
185	185
197	-
210	211
226	229

3.1.3 Mutante - D85S

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1JV7 [RCF⁺01] zeichnet sich durch eine D85S Mutation aus. Die Aminosäure Asparaginsäure an der Position 85 wurde dabei durch die Aminosäure Serin ausgetauscht. Diese Mutation führt zu einer Neutralisation des Schiffschens Basen Counterions bei Aminosäure 85. Des Weiteren führt diese Mutation zu einigen Veränderungen in den extrazellulären Bereichen des Proteins.

Tabelle 3.3: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	60
72	73
88	84
105/116	111
140	138/144
154	154
185	185
197	-
210	210
226	221/229

3.1.4 Mutante - F219L

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1P8I [[SBL03](#)] zeichnet sich durch eine F219L Mutation aus. Die Aminosäure Phenylalanin an der Position 219 wurde dabei durch die Aminosäure Leucin ausgetauscht. Diese Mutation führt zu einer Ausbuchtung beim Aminosäurerest 219 in dem sich drei Wassermoleküle anlagerten.

Tabelle 3.4: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	60
72	73
88	84
105/116	111
140	138/144
154	154
185	185
197	-
210	210
226	221/229

3.1.5 Mutante - P50A

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1PXR [FYB⁺04] zeichnet sich durch eine P50A Mutation aus. Die Aminosäure Prolin an der Position 219 wurde dabei durch die Aminosäure Alanin ausgetauscht. Diese Mutation hat keinerlei Auswirkungen auf die Stabilität des Proteins.

Tabelle 3.5: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	9
17/24	20
30	-
48	51
60	62
72	72
88	84
105/116	110
140	144
154	-
-	162
185	185
197	-
210	210
226	-

3.1.6 Mutante - M56A

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1PXS [FYB⁺04] zeichnet sich durch eine M56A Mutation aus. Die Aminosäure Methionin an der Position 56 wurde dabei durch die Aminosäure Alanin ausgetauscht.

Tabelle 3.6: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17	18
24	22
30	33
48	53
60	62
72	72
88	84
105/116	111
-	131
140	142
154	-
185	185
197	198
210	209
226	221

3.1.7 Mutante - P186A

Dieser Mutant des Proteins Bacteriorhodopsin mit der PDB-ID 1Q5I [YFY⁺04b] zeichnet sich durch eine P186A Mutation aus. Die Aminosäure Prolin an der Position 186 wurde dabei durch die Aminosäure Alanin ausgetauscht. Diese Mutation hat fast keinen Einfluss auf die Bildung der transmembranen Helix und somit auf die Gesamtstruktur. Diese Mutation ist eine von drei, bei der Proline in den transmembranen Helices gegen Alanine ausgetauscht werden, ohne das Struktur oder Funktion verloren gehen.

Tabelle 3.7: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17	15
24	24
30	33
48	54
60	60
72	72
88	84
-	95
105	108
116	120
-	131
140	142
154	-
185	184
197	-
210	207
-	218
226	228

3.1.8 Mutante - P91A

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1Q5J [YFY⁺04b] zeichnet sich durch eine P91A Mutation aus. Die Aminosäure Prolin an der Position 91 wurde dabei durch die Aminosäure Alanin ausgetauscht. Wie bei der Mutante P186A ist diese Mutation eine der drei Proline in transmembranen Helices, welche ohne Verlust der Struktur und Funktion gegen Alanine ausgetauscht werden können.

Tabelle 3.8: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17	18
24	22
30	33
48	48
60	62
72	72
88	84
105	109
116	113
140	140
154	154
185	185
197	-
210	207

3.1.9 Mutante - T24S

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1S51 [YFY⁺04a] zeichnet sich durch eine T24S Mutation aus. Das bedeutet, dass die Aminosäure Threonin an der Position 24 wurde dabei durch die Aminosäure Serin ausgetauscht. Diese Mutante zeigt keine auffälligen strukturellen Veränderungen und ist nur geringfügig instabiler als der Wildtyp. Dies wird bedingt durch die Auflösung der C_{α} -Wasserstoffbrücke zum strukturell benachbarten Aminosäurerest 51.

Tabelle 3.9: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17	18
24	22
30	33
48	53
60	63
72	72
88	84
105	109
116	1113
-	131
140	142
154	154
185	185
197	198
210	207
226	221

3.1.10 Mutante - T24V

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1S52 [YFY⁺04a] zeichnet sich durch eine T24V Mutation aus. Die Aminosäure Threonin an der Position 24 wurde dabei durch die Aminosäure Valin ausgetauscht. Diese Mutante zeigt keine auffälligen strukturellen Veränderungen und ist nur geringfügig instabiler als der Wildtyp. Dies wird bedingt durch die Auflösung der C_{α} -Wasserstoffbrücke zum strukturell benachbarten Aminosäurerest 51.

Tabelle 3.10: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17/24	20
30	33
48	54
60	60
72	73
88	84
105/116	111
140	138/144
154	-
185	185
197	-
210	207
226	221

3.1.11 Mutante - T24A

Diese Mutante des Proteins Bacteriorhodopsin mit der PDB-ID 1S54 [YFY⁺04a] zeichnet sich durch eine T24A Mutation aus. Die Aminosäure Threonin an der Position 24 wurde dabei durch die Aminosäure Alanin ausgetauscht. Diese Mutante zeigt keine auffälligen strukturellen Veränderungen und ist sogar stabiler als der Wildtyp.

Tabelle 3.11: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	6
17/24	20
30	-
48	-
60	62
72	72
88	84
105/116	111
140	138/144
154	154
185	185
197	-
210	207
226	-

3.1.12 Mutante - G113L

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine G113L Mutation aus. Die Aminosäure Glycin an der Position 113 wurde dabei durch die Aminosäure Leucin ausgetauscht.

Tabelle 3.12: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	60
72	73
88	86
105/116	110
140	138/144
154	-
185	185
197	-
210	210
226	221

3.1.13 Mutante - G113Q

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine G113Q Mutation aus. Die Aminosäure Glycin an der Position 113 wurde dabei durch die Aminosäure Glutamin ausgetauscht.

Tabelle 3.13: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	53
60	63
72	73
88	86
105	108
116	-
-	132
140	145
154	-
185	185
197	-
210	210
226	229

3.1.14 Mutante - G116L

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine G116L Mutation aus. Die Aminosäure Glycin an der Position 116 wurde dabei durch die Aminosäure Leucin ausgetauscht.

Tabelle 3.14: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	60
72	73
88	84
105/116	111
140	138/144
154	154
185	185
197	-
210	210
226	221/229

3.1.15 Mutante - G116Q

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine G116Q Mutation aus. Die Aminosäure Glycin an der Position 113 wurde dabei durch die Aminosäure Glutamin ausgetauscht.

Tabelle 3.15: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	53
60	62
72	73
88	84
105	111
140	140
154	154
185	185
197	-
210	210
226	221/229

3.1.16 Mutante - I117A

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine I117A Mutation aus. Die Aminosäure Isoleucin an der Position 117 wurde dabei durch die Aminosäure Alanin ausgetauscht.

Tabelle 3.16: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	60
72	73
88	84
105	-
116	115
140	138/144
154	154
185	185
197	-
210	210
226	221/229

3.1.17 Mutante - I117F

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine I117F Mutation aus. Die Aminosäure Isoleucin an der Position 117 wurde dabei durch die Aminosäure Phenylalanin ausgetauscht.

Tabelle 3.17: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	-
60	62
72	73
88	86
105	110
116	-
140	138/144
154	-
185	185
197	-
210	210
226	221/229

3.1.18 Mutante - M145F

Diese Mutante des Proteins Bacteriorhodopsin zeichnet sich durch eine M145F Mutation aus. Die Aminosäure Methionin an der Position 145 wurde dabei durch die Aminosäure Phenylalanin ausgetauscht.

Tabelle 3.18: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
10	10
17	17
24	24
30	30
48	53
60	62
72	73
88	84
105	111
116	-
- 140	140
154	154
185	185
197	-
210	210
226	221/229

3.2 Halorhodopsin

Die Ergebnisse der Barrierenvorhersage werden direkt mit im Experiment ermittelten Barrieren verglichen, wobei das Fenster zur Bestimmung von richtig Vorhergesagten Entfaltungsbarrieren beachtet werden muss (vgl. Kapitel 2.2) [COM05].

Tabelle 3.19: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein. In der rechten Spalte sind die experimentellen Daten aufgetragen. Ein ' - ' in einer Spalte bedeutet, dass es für diesen Wert keine Vorhersage im Bereich des Fensters (vgl. Kapitel 2.2) oder keine experimentellen Daten gibt. Ein ' / ' in einer Spalte bedeutet, dass mehrere Daten miteinander vereinbar sind.

Vorhersage Halorhodopsin	experimentelle Daten
41	37
53	51
66	-
83	-
93	94
113	114
132	127
141	-
156	-
169	171
180	177
191	186/190
222	227
236	237
244	248

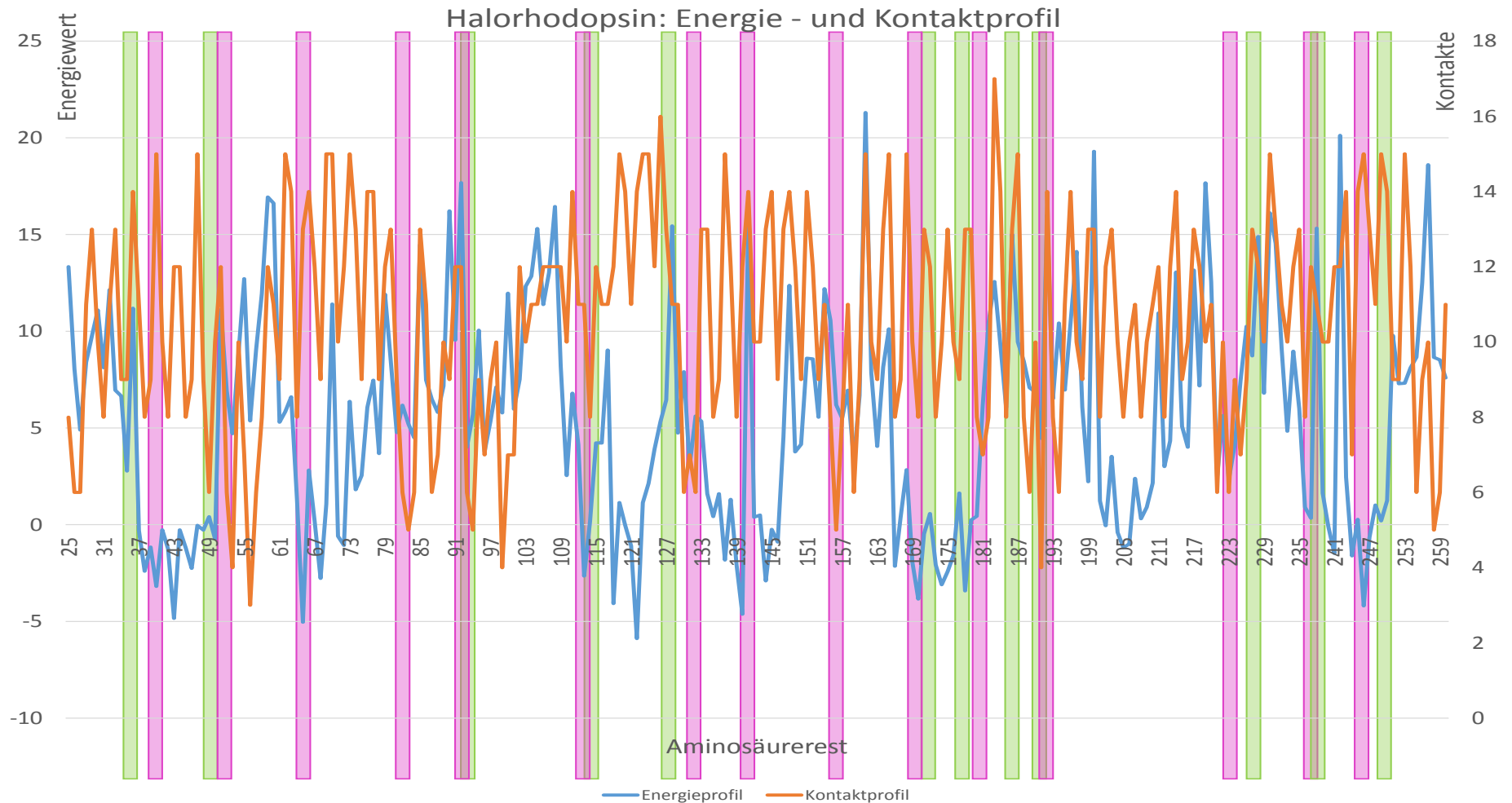


Abbildung 3.2: Darstellung der Ergebnisse der Barrierenvorhersage des Proteins Halorhodopsin. In blau ist der Verlauf der Energiewerte zum jeweiligen Aminosäurerest gegen die linke y-Achse aufgetragen. In braun ist Verlauf der Kontaktanzahl zum jeweiligen Aminosäurerest gegen die rechte y-Achse aufgetragen. Die grünen Balken zeigen die experimentell ermittelten Barrieren, während die magentafarbenen Balken die vorhergesagten Aminosäurereste anzeigen. Grüne Balken mit magentafarbenem Rand zeigen eine genaue Übereinstimmung von Vorhersage und Experiment.

3.2.1 Mutationen

Für das Protein Halorhodopsin wurden insgesamt vier verschiedene Mutationen betrachtet, um den Einfluss dieser auf die Vorhersagen zu untersuchen. Diese Betrachtung ist notwendig, um einerseits den Einflussbereich der Mutation überprüfen zu können und um andererseits die Möglichkeit zu prüfen, ob es möglich ist, mit nur einer Mutation eine Barrierenvorhersage zu verursachen bzw. eine Vorhersage ausfallen zu lassen. Hierzu wurden sowohl aufgeklärte als auch modellierte Proteine betrachtet.

Um die Daten optimal vergleichen zu können, werden die Vorhersagen der Mutanten mit denen des nativen Proteins verglichen.

3.2.2 Mutante - H95A

Diese Mutante des Proteins Halorhodopsin zeichnet sich durch eine H95A Mutation aus. Die Aminosäure Histidin an der Position 95 wurde dabei durch die Aminosäure Alanin ausgetauscht.

Tabelle 3.20: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
-	22
41	-
53	52
66	-
-	74
83	-
93	95
113	110
132	125
141	135
156	150
169	-
180	183
191	-
-	202
-	212
222	224
236	237
244	-

3.2.3 Mutante - H95R

Diese Mutante des Proteins Halorhodopsin zeichnet sich durch eine H95R Mutation aus. Die Aminosäure Histidin an der Position 95 wurde dabei durch die Aminosäure Arginin ausgetauscht.

Tabelle 3.21: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
-	22
41	-
53	52
66	-
-	73
83	-
93	95
113	110
132	125
141	135
156	150
169	-
180	183
191	-
-	202
-	212
222	224
236	237
244	-

3.2.4 Mutante - R108Q

Diese Mutante des Proteins Halorhodopsin zeichnet sich durch eine R108Q Mutation aus. Die Aminosäure Arginin an der Position 108 wurde dabei durch die Aminosäure Glutamin ausgetauscht.

Tabelle 3.22: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
-	22
41	-
53	52
66	-
-	74
83	-
93	95
113	110
132	125
141	135
156	150
169	-
180	183
191	-
-	202
-	212
222	224
236	237
244	-

3.2.5 Mutante - T203V

Diese Mutante des Proteins Halorhodopsin zeichnet sich durch eine T203V Mutation aus. Die Aminosäure Threonin an der Position 203 wurde dabei durch die Aminosäure Valin ausgetauscht.

Tabelle 3.23: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein und in der rechten Spalte sind die Vorhersagen des Mutanten aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen des anderen Proteins vereinbar sind.

Vorhersage natives Protein	Vorhersage Mutant
-	22
41	-
53	52
66	-
-	74
83	-
93	95
113	110
132	125
141	135
156	150
169	-
180	183
191	-
-	202
-	212
222	224
236	237
244	-

3.3 Humane Aquaporine 1

Beim humanen Aquaporin 1 wird, im Gegensatz zu den anderen betrachteten Proteinen, nicht nur das native Protein betrachtet und verglichen, sondern auch die bekannten homologen humanen Aquaporine 2 bis 11 eingeführt. Des Weiteren werden die Ergebnisse von drei unterschiedlich aufgelösten humanen Aquaporinen 1 direkt mit den experimentellen Daten verglichen [MFS⁺03].

3.3.1 Vergleich der aufgeklärten humanen Aquaporine 1

In Tabelle 3.24 ist der Vergleich der Barrierenvorhersagen der drei in unterschiedlicher Auflösung bekannten humanen Aquaporine 1 aufgetragen (Stand Juni 2013). Diese Proteine haben folgende PDB-ID's: 1FQY [MMH⁺00], 1H6I [dEG01] und 1IH5 [RRC⁺01]. Die Vorhersagen werden direkt mit den experimentellen Daten verglichen, wobei das Fenster zur Bestimmung von richtig vorhergesagten Entfaltungsbarrieren beachtet werden muss (vgl. Kapitel 2.2).

Tabelle 3.24: In der ersten Spalte steht der vorhergesagte Aminosäurerest für das Protein mit der PDB-ID 1FQY, in der zweiten Spalte steht der vorhergesagte Aminosäurerest für das Protein mit der PDB-ID 1H6I, in der dritten Spalte steht der vorhergesagte Aminosäurerest für das Protein mit der PDB-ID 1IH5 und in der rechten Spalte sind die experimentell ermittelten Barrieren aufgetragen. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich des anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen der anderen Proteine vereinbar sind.

Vorhersage 1FQY	Vorhersage 1H6I	Vorhersage 1IH5	experimentelle Daten
23	20	-	25
38	43	38	43
48	55	-	-
68	65	62	64
-	75	-	80
88	-	89	85
-	93	-	-
108	110	102/108	-
120	121	121	124
136	131	133	132
-	142	-	-
-	153	-	-
160	161	164	162
185	179	182	184
206	204	206	205
219	221	217	217
230			

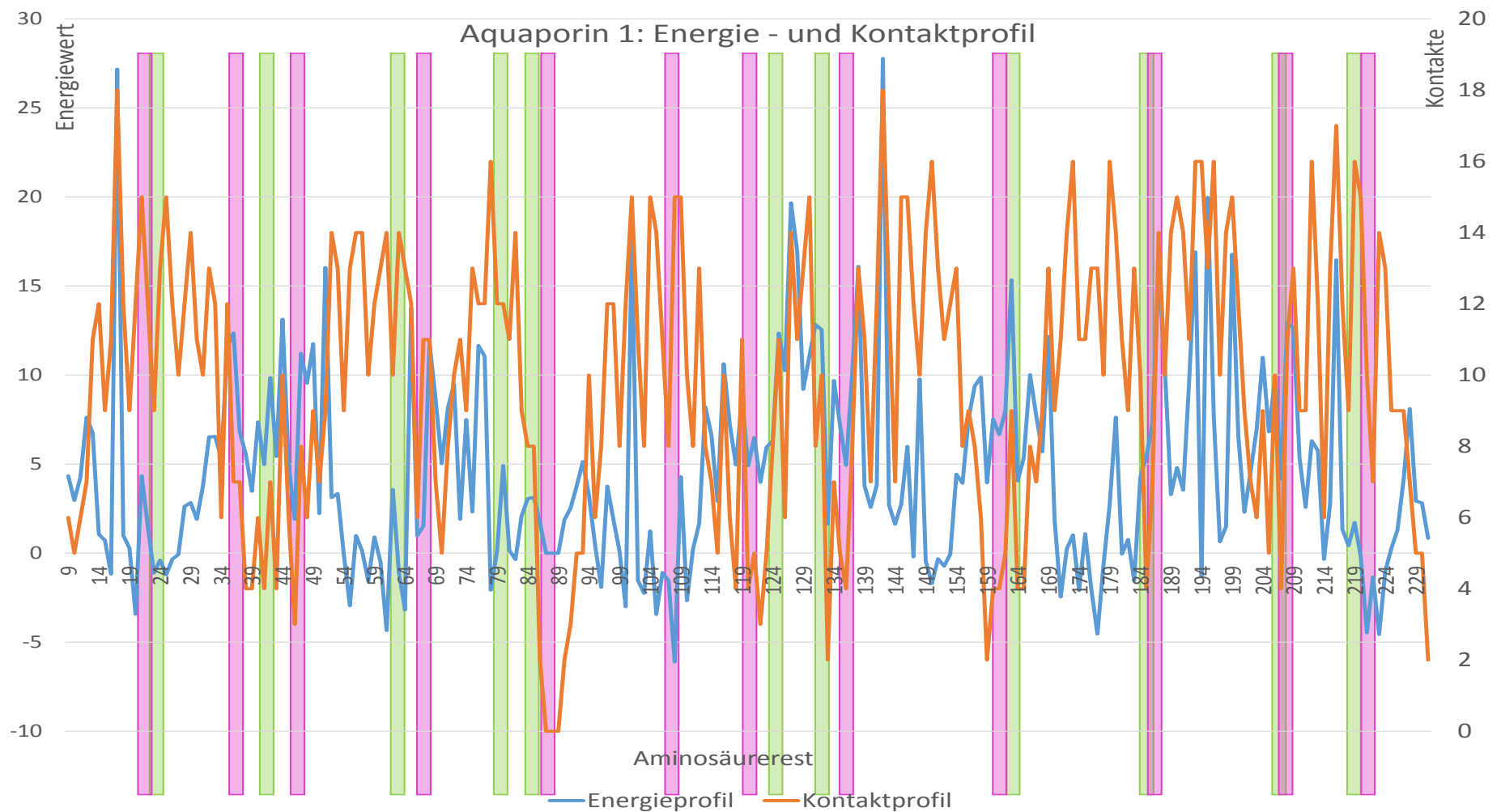


Abbildung 3.3: Darstellung der Ergebnisse der Barrierenvorhersage des Proteins Aquaporin 1 mit der PDB-ID 1FQY. In blau ist der Verlauf der Energiewerte zum jeweiligen Aminosäurerest gegen die linke y-Achse aufgetragen. In braun ist Verlauf der Kontaktanzahl zum jeweiligen Aminosäurerest gegen die rechte y-Achse aufgetragen. Die grünen Balken zeigen die experimentell ermittelten Barrieren, während die magentafarbenen Balken die vorhergesagten Aminosäurereste anzeigen. Grüne Balken mit magentafarbenem Rand zeigen eine genaue Übereinstimmung von Vorhersage und Experiment.

3.3.2 Vergleich der vorhergesagten Barrieren von Aquaporin 1 bis 11

In diesem Abschnitt werden die Ergebnisse der Vorhersagen der einzelnen Aquaporine tabellarisch dargestellt. Hierbei kommt es sowohl zur Verwendung von aufgeklärten als auch von vorhergesagten Proteinstrukturen. Die Aquaporine mit aufgeklärter Struktur sind: Aquaporin 1 - 1FQY, Aquaporin 4 - 3GD8 [HYS⁺09] , Aquaporin 5 - 3D9S [HNF⁺08]. Die Strukturen von Aquaporin 2 ,3 ,6 ,7 ,8 ,9 ,10 wurden der ModBase entnommen, da hier eine gute Genauigkeit der Strukturvorhersage gegeben war, lediglich Aquaporin 11 wurde mit SwissModel modelliert und hatte eine höhere Güte, als das Modell von ModBase.

Tabelle 3.25: In der linken Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 1 mit der PDB-ID 1FQY, in der zweiten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 2, in der dritten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 3, in der vierten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 4 mit der PDB-ID 3GD8, in der fünften Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 5 mit der PDB-ID 3D9S und in der rechten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 6. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich der anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen der anderen Proteine vereinbar sind.

Vorhersage AQ1	Vorhersage AQ2	Vorhersage AQ3	Vorhersage AQ4	Vorhersage AQ5	Vorhersage AQ6
23	15/24	26	-	17	23
38	39	45	34	38	34
48	55	45	45	47/59	50
68	73	64/75	73	78	69
88	81	89	87	88	-
-	-	98	-	102	102
108	101/107	107	114	108	102
120	125	117	116	123	118
136	136	131	131	131	137
-	146	144	144	147	146
160	154	152/160	152/165	155	-
-	171	173	173	171	
185	192	185	182	196	182
206	208	203	200	204	212
219	221	-	225	217	220
230	231	229	236	225/233	238
-	-	246	245	243	238
-	-	258	-	-	252

Tabelle 3.26: In der linken Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 7, in der zweiten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 8, in der dritten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 9, in der vierten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 10 und in der rechten Spalte steht der vorhergesagte Aminosäurerest für Aquaporin 11. Dabei bedeutet ein ' - ' in einer Spalte, dass es für diesen Wert keine Vorhersage im Bereich der anderen Proteins gibt. Ein ' / ' in einer Spalte bedeutet, dass die vorhergesagten Daten mit mehreren Vorhersagen der anderen Proteine vereinbar sind.

Vorhersage AQ7	Vorhersage AQ8	Vorhersage AQ9	Vorhersage AQ10	Vorhersage AQ11
-	-	31	31	-
48	45	41	-	43
56	53	54	52	52
64	61/70	73	71	72
87	79/88	90	88	86
106	105	103	101	100
113	117	118	110	114
120	117	118	124	125
129	128	135	133	134
144	147	141	146	151
153	147	154	-	151
163	170	165	163	171
171	170	173	-	- 171
189	185/192	186	185/194	188
204	198	204	205	204
216	222	-	-	-
-	230	229	227	225
240	238	243	241	-
250	253	251	249	249
261	-	260	258	-

3.4 Natives Rinderrhodopsin

Die Ergebnisse der Barrierenvorhersage werden direkt mit im Experiment ermittelten Barrieren verglichen, wobei das Fenster zur Bestimmung von richtig vorhergesagten Entfaltungsbarrieren beachtet werden muss (vgl. Kapitel 2.2) [TPF⁺06].

Tabelle 3.27: In der linken Spalte steht der vorhergesagte Aminosäurerest für das native Protein. In der rechten Spalte sind die experimentellen Daten aufgetragen. Ein ' - ' in einer Spalte bedeutet, dass es für diesen Wert keine Vorhersage im Bereich des Fensters (vgl. Kapitel 2.2) oder keine experimentellen Daten gibt. Ein ' / ' in einer Spalte bedeutet, dass mehrere Daten miteinander vereinbar sind.

Vorhersage natives Rinderhodopsin	experimentelle Daten
17	19
28	26
41	35
48/52	45
63	64
86	82
99	97
108	108
129	123
141	-
155	159
164/173	169
197/206	199
219/225	219
236	243
262/269	262
275	-
286	-
300	296
-	314
323	-

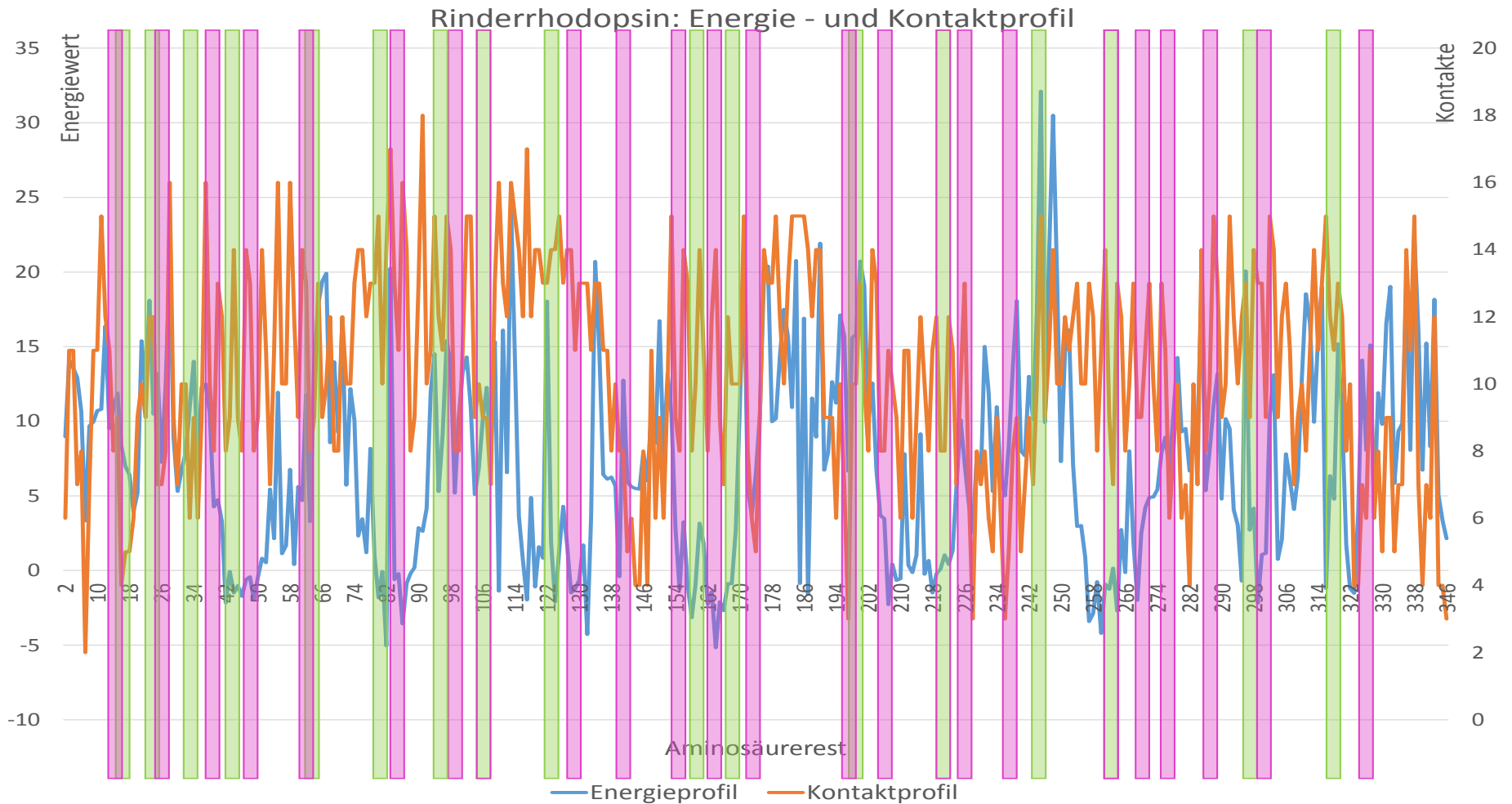


Abbildung 3.4: Darstellung der Ergebnisse der Barrierenvorhersage des Proteins Rinderrhodopsin. In blau ist der Verlauf der Energiewerte zum jeweiligen Aminosäurerest gegen die linke y-Achse aufgetragen. In braun ist Verlauf der Kontaktanzahl zum jeweiligen Aminosäurerest gegen die rechte y-Achse aufgetragen. Die grünen Balken zeigen die experimentell ermittelten Barrieren, während die magentafarbenen Balken die vorhergesagten Aminosäurereste anzeigen. Grüne Balken mit magentafarbenem Rand zeigen eine genaue Übereinstimmung von Vorhersage und Experiment.

3.5 Zusammenfassung der Ergebnisse

In der folgenden Abbildung 3.5 sind die Ergebnisse der Vorhersagen der Entfaltungsbarrieren der einzelnen Proteine schematisch gegen die Struktur des betreffenden Proteins aufgetragen.

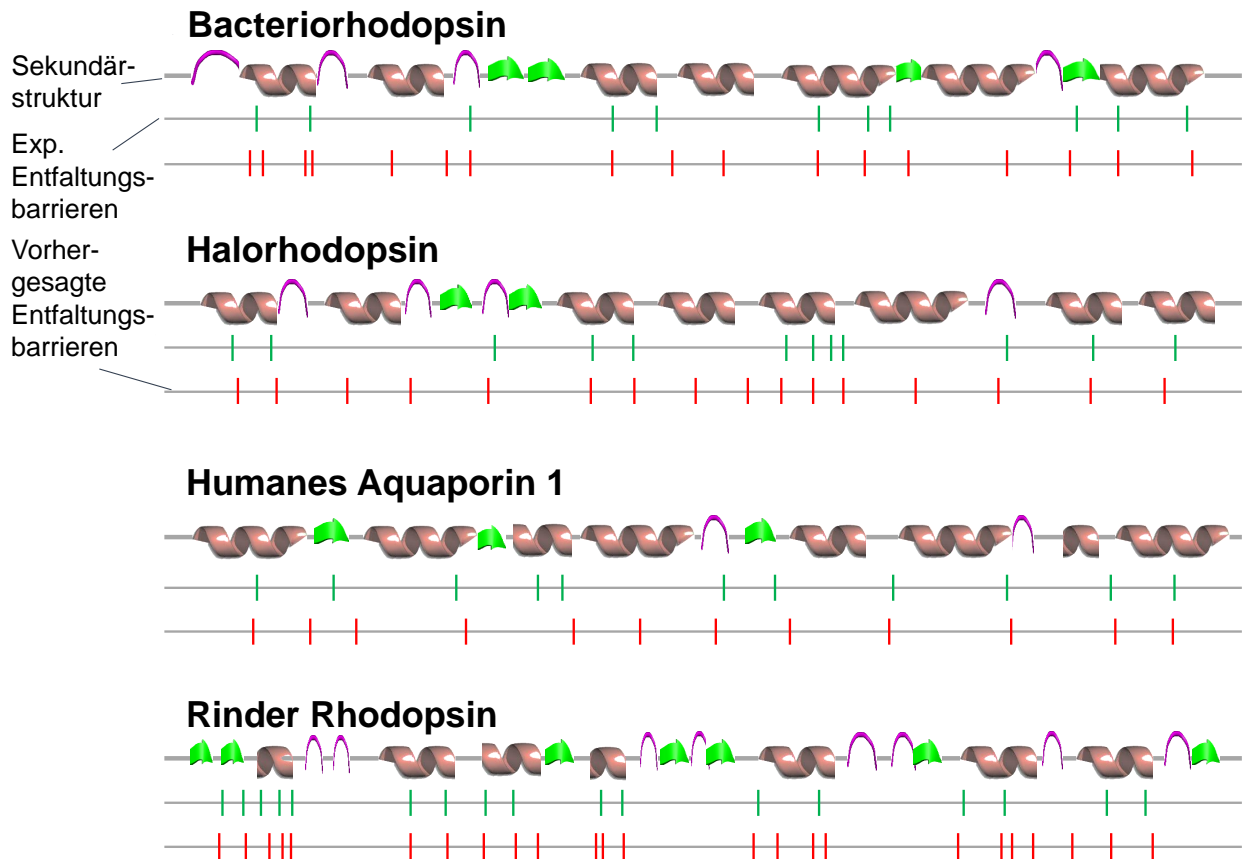


Abbildung 3.5: Vergleich der vorhergesagten Entfaltungsbarrieren mit denen im Experiment ermittelten aufgetragen gegen eine schematische Struktur des betreffenden Proteins. Hierbei stellt die obere Spalte jeweils die schematische Sekundärstruktur des Proteins nach der UniprotKB dar. Die zweite Zeile zeigt die experimentell ermittelten Entfaltungsbarrieren und die dritte Zeile zeigt die vorhergesagten Entfaltungsbarrieren. Da es sich um eine schematische Abbildung handelt, können die dargestellten Abstände zu den reellen Abständen unterschiedlich sein.

4 Evaluation

In diesem Kapitel werden die Ergebnisse gesammelt und mit den in Kapitel 1.5 vorgestellten Gütemaßen evaluiert. Hierbei werden nicht nur die Ergebnisse aus dieser Arbeit, welche nur mit den nativen Proteinen arbeitet vorgestellt, sondern auch die Ergebnisse, welchen die Simulation des Experiments zu Grunde liegt, um die Verwendung des nativen Zustandes besser nachvollziehen zu können.

4.1 Mit Experiment-Simulation

Die Ergebnisse, welche mit einer Simulation des Experiments erzielt wurden, werden nicht explizit, sondern lediglich in zusammengefasster Form vorgestellt. Somit ergibt sich für diese Ergebnisse folgende Tabelle:

Tabelle 4.1: Darstellung der Ergebnisse für die Simulation des Experiments

		PDB ID's der Proteine				
		1BRR	1E12	1FQY	1U19	insgesamt
Vorhersage	Anzahl AS	226	235	222	345	1028
	TP	11	12	11	17	51
	TN	209	217	203	320	949
	FN	0	0	0	0	0
	FP	6	6	8	8	28

Aus dieser Tabelle ergibt sich folgende Konfusionsmatrix:

Tabelle 4.2: Konfusionsmatrix für die Simulation des Experiments

		Experiment	
		positiv	negativ
Vorhersage	positiv	51	28
	negativ	0	949

Anhand dieser Ergebnisse werden die in Kapitel 1.5 vorgestellten Gütemaße für dieses Verfahren berechnet. Für den MCC ergibt sich folgendes Ergebnis analog zur Gleichung 1.7 :

$$(4.1) \quad MCC = \frac{51 \times 949 - 28 \times 0}{\sqrt{(51 + 28) \times (51 + 0) \times (949 + 0) \times (949 + 28)}} = 0,792$$

Um nun das F1-Maß berechnen zu können sind zunächst die in Kapitel 1.5 eingeführten Größen Genauigkeit und Trefferquote analog zu den Gleichungen 1.8 und 1.9, zu ermitteln.

$$(4.2) \quad \text{Genauigkeit} = \frac{51}{(51 + 28)} = 0,646$$

$$(4.3) \quad \text{Trefferquote} = \frac{51}{(51 + 0)} = 1,0$$

Mit Hilfe dieser beiden Größen berechnet man nun analog zur Gleichung 1.10 das F1-Maß:

$$(4.4) \quad F1 = 2 \times \frac{(0,646 \times 1,0)}{(0,646 + 1,0)} = 0,785$$

4.2 Nativer Zustand

Dieser Abschnitt dient der Darstellung der Ergebnisse, welche explizit in den einzelnen Abschnitten des Kapitels 3 dargestellt wurden, und deren Evaluierung.

Aus den dargestellten Tabellen in Kapitel 3, welche die Vorhersagen für Entfaltungsbarrrieren für die einzelnen nativen Proteine enthalten, ergibt sich, analog zum vorigen Abschnitt 4.1, folgende Tabelle:

Tabelle 4.3: Darstellung der Ergebnisse für den nativen Zustand

		PDB ID's der Proteine				
		1BRR	1E12	1FQY	1U19	insgesamt
Vorhersage	Anzahl AS	226	235	222	345	1032
	TP	11	12	11	16	50
	TN	211	218	210	324	967
	FN	0	0	0	1	1
	FP	4	5	1	4	14

Aus dieser Tabelle ergibt folgende Konfusionsmatrix:

Tabelle 4.4: Konfusionsmatrix für die Simulation des Experiments

		Experiment	
		positiv	negativ
Vorhersage	positiv	50	14
	negativ	1	967

Anhand dieser Ergebnisse werden die in Kapitel 1.5 vorgestellten Gütemaße für dieses Verfahren berechnet. Für den MCC ergibt sich folgendes :

$$(4.5) \quad MCC = \frac{50 \times 967 - 14 \times 1}{\sqrt{(50 + 14) \times (50 + 1) \times (967 + 1) \times (967 + 14)}} = 0,868$$

Zunächst sind nun die in Kapitel 1.5 eingeführten Größen Genauigkeit und Trefferquote zu berechnen.

$$(4.6) \quad \text{Genauigkeit} = \frac{50}{(50 + 14)} = 0,781$$

$$(4.7) \quad \text{Trefferquote} = \frac{50}{(50 + 1)} = 0,980$$

Mit Hilfe dieser beiden Größen berechnet man nun das F1-Maß:

$$(4.8) \quad F1 = 2 \times \frac{(0,781 \times 0,980)}{(0,781 + 0,980)} = 0,869$$

4.3 Signifikanztest mittels randomisierter Daten

Ein weiteres Verfahren, um die Güte eines Algorithmus zu bestärken, ist der Vergleich zu randomisierten Daten. Es wurde kein komplett auf Zufall basierendes Verfahren verwendet, sondern ein auf realen Werten basierendes. Das bedeutet im Speziellen, dass die Erzeugung der Binärsequenz, wie in Kapitel 2.1 bereits eingeführt, nicht gleichverteilt zufällig ist, sondern auf speziellen Werten beruht. Um diese speziellen Werte

zu erhalten, wurden die vor der ersten Zusammenfassung erhaltenen Binärsequenzen verwendet. Von diesen Sequenzen wurde nun ermittelt, wie oft auf eine 0 (keine Vorhersage) eine 0, auf eine 0 eine 1 (Vorhersage), auf eine 1 eine 0 und auf eine 1 eine weitere 1 folgt.

Aus den vier betrachteten Proteinen ergibt sich folgende Tabelle, in der die Anzahl der aufeinander folgenden Werte dargestellt ist.

Tabelle 4.5: Darstellung der Anzahl, wie häufig auf eine 0 (keine Vorhersage) eine 0, auf eine 0 eine 1 (Vorhersage), auf eine 1 eine 0 und auf eine 1 eine weitere 1 folgt.

	0	1
0	614	165
1	168	88

Mit der Gesamtanzahl der Vorhersagen lassen sich mit Hilfe der in Tabelle 4.5 erhaltenen Werte Wahrscheinlichkeiten berechnen.

Tabelle 4.6: Darstellung der Prozentwerte, mit welchen Wahrscheinlichkeiten auf eine 0 (keine Vorhersage) eine 0, auf eine 0 eine 1 (Vorhersage), auf eine 1 eine 0 und auf eine 1 eine weitere 1 folgt.

	0	1
0	78.82	21.18
1	65.62	34.38

Mit Hilfe dieser Werte wurden nun 200001 Vorhersagen gemacht, bei denen mit 78.82%-iger Wahrscheinlichkeit auf eine nicht Vorhersage eine nicht Vorhersage, mit 21.18%-iger Wahrscheinlichkeit auf eine nicht Vorhersage eine Vorhersage, mit 65.62%-iger Wahrscheinlichkeit auf eine Vorhersage eine nicht Vorhersage und mit 34.38%-iger Wahrscheinlichkeit auf eine Vorhersage eine weitere Vorhersage folgt.

Alle zufällig erzeugten Vorhersagesequenzen wurden anschließend, genau wie bei der Betrachtung der anderen Proteine, zusammengefasst und evaluiert (vgl. Abbildung 2.3 und 2.4).

Für alle 200001 Zufallsvorhersagen ergibt sich ein durchschnittlicher MCC von 0,691 mit einer Standardabweichung von 0,0299 und einem durchschnittlichen F1 von 0,694 mit einer Standardabweichung von 0,0297.

Somit ist, ausgehend von einem MCC von 0,868 und einem F1-Maß von 0.869 für die Betrachtung des nativen Zustands der Proteine, der entwickelte und verwendete Vorhersagealgorithmus, auf den MCC bezogen, 5.919 Standardabweichungen, und auf

das F1-Maß bezogen, 5,892 Standardabweichungen mal besser, als die erzeugte Vorhersagen, welche auf randomisierten Daten beruhen und damit die selben zugrunde liegenden Wahrscheinlichkeiten haben.

Bei einer komplett zufälligen Vorhersage, bei der alle Werte der Wahrscheinlichkeitstabelle 50% betragen, ist das Ergebnis nur wenig unterschiedlich zu dem bisher genannten.

Auch hier wurden 200001 Vorhersagen durchgeführt und liefern einen durchschnittlichen MCC von 0,665 mit einer Standardabweichung von 0,0155 und einem durchschnittlichen F1 von 0,658 mit einer Standardabweichung von 0,0176.

Diese Vorhersage ist somit 0,026 bzw. 0,036 schlechter als die auf berechneten Wahrscheinlichkeiten beruhende. Dadurch, dass die Standardabweichungen kleiner sind, kann bei der zufälligen Vorhersage von konstanten Ergebnissen ausgegangen werden.

5 Diskussion

Dieser Abschnitt befasst sich mit der Diskussion der in der Arbeit erzielten Ergebnisse und Erkenntnisse.

5.1 Mutationen

Im folgenden werden die Ergebnisse der Betrachtung der verschiedenen Mutationen gesammelt und mit den in der Literatur bekannten Effekten verglichen. [SBL⁺08] Die Tabelle wurde adaptiert [Dre08] und mit den neuen Erkenntnissen ergänzt bzw. verbessert.

Um den globalen Einfluss der Mutationen geltend zu machen und sie nicht nur punktuell zu betrachten, wird für die Aussage in der folgenden Tabelle 5.1 das Gesamtpotential der Energien der Aminosäurereste berechnet. Ist das Gesamtpotential der Mutation höher als der des Wildtyps, sprechen wir von einer destabilisierenden, bei einer Erniedrigung des Gesamtpotentials gegenüber des Wildtyps von einer stabilisierenden Mutation.

Tabelle 5.1: Tabellarische Darstellung der Ergebnisse der Einflüsse der Mutationen auf die Stabilität des Proteins. Hierbei bedeutet ein 'l.' leicht und ein '-', dass keine vergleichende Aussage getroffen werden konnte.

Protein	Mutation	Auswirkung in der Literatur	Vorhersage	Kongruenz
Bacteriorhodopsin	D96N	destabilisierend	stabilisierend	nein
	D85S	keine Aussage	stabilisierend	-
	F219L	keine Aussage	l. stabilisierend	-
	P50A	keine Auswirkung	stabilisierend	nein
	M56A	keine Auswirkung	destabilisierend	nein
	P186A	keine Auswirkung	stabilisierend	nein
	P91A	keine Auswirkung	stabilisierend	nein
	T24S	destabilisierend	destabilisierend	ja
	T24V	destabilisierend	destabilisierend	ja
	T24A	destabilisierend	destabilisierend	ja
	G113L	destabilisierend	l. stabilisierend	nein
	G113Q	destabilisierend	destabilisierend	ja
	G116L	destabilisierend	l. stabilisierend	nein
	G116Q	destabilisierend	destabilisierend	ja
	I117A	destabilisierend	l. stabilisierend	nein
	I117F	destabilisierend	l. stabilisierend	nein
	M145F	aktiv	keine Auswirkung	ja
Halorhodopsin	H95A	destabilisierend	stabilisierend	nein
	H95R	destabilisierend	stabilisierend	nein
	R108Q	inaktiv	stabilisierend	-
	T203V	eingeschränkt aktiv	stabilisierend	-

Anhand dieser Tabelle kann man erkennen, dass die in der Literatur vorhandene Aussage über die Auswirkung auf die Stabilität einer Mutation keinesfalls mit dem hier verwendeten Energieprofil einhergehen muss. Genau genommen stimmen lediglich sechs der 20 Aussagen mit der Literatur überein. Das entspricht 30%, wobei zu 4 Mutationen keine vergleichende Aussage getroffen werden kann. Dies ist bedingt dadurch, dass das hier verwendete und ausgewertete Energieprofil keinerlei Angaben zur Aktivität oder Inaktivität eines Proteins geben kann. Hierbei muss man beachten, dass auch wenn z. B. in der Literatur eine Mutation ohne Effekt angegeben wird, sie dennoch Einfluss auf die Stabilität nehmen kann, welche allerdings keine Auswirkungen auf die Aktivität haben. Die in dieser Arbeit untersuchten Mutationen zeigen klar und deutlich, dass die Betrachtung einer einzigen Mutation mit dem Energieprofil eine Aussage über die Stabilität des Gesamtproteins zulassen. Hierbei gilt weiterhin zu beachten, dass die Effekte, welche in der oben stehenden Tabelle 5.1 genannt werden, sich ausschließlich auf die Betrachtung des Energieprofils beziehen, nicht auf das Kontaktprofil bzw. beide Profile. Hierbei ist, wie bei allen anderen unterschiedlich aufgeklärten Proteinen auffällig, dass sich das

Kontaktprofil der Proteine erheblich unterscheidet. Dies hat, verglichen mit der Berechnung des Energieprofils aus Kapitel 1.3.2, einen erheblichen Einfluss auf eben dieses, wenn, aufgrund der Aufklärung oder der Mutation, spezifische Kontakte einer Aminosäure wegfallen. Es lässt sich im Rahmen der hier getätigten Vorhersage allerdings nicht exakt nachprüfen, ob ein Kontakt aufgrund der Aufklärung oder seiner Strukturvorhersage wegfällt oder ob es an veränderten sterischen Begebenheiten in dem Protein liegt. Da es sich bei den betrachteten Proteinen um Tunnelproteine handelt, können diese sterischen Begebenheiten ebenfalls durch den unterschiedlichen Zustand in dem sich das Protein zur Zeit seiner Aufklärung, befindet, entstehen. Die folgenden vier Abbildungen sollen exemplarisch die Unterschiede in den Kontakt- und Energieprofilen bei Mutationen im Vergleich zum Wildtyp zeigen. Die Abbildungen 5.1 und 5.2 zeigen dabei exemplarisch die Unterschiede in den Profilen von Bacteriorhodopsin zu einer Mutanten, während die Abbildung 5.3 und 5.4 die Unterschiede von Halorhodopsin zu einer seiner Mutanten verdeutlichen sollen. Die betrachtete Mutante für Bacteriorhodopsin ist dabei die zuvor betrachtete D96N Mutante mit der PDB-ID 1C8R (vgl. Kapitel 3.1.2), während bei Halorhodopsin die H95A Mutante (vgl. Kapitel 3.2.2) betrachtet wird.

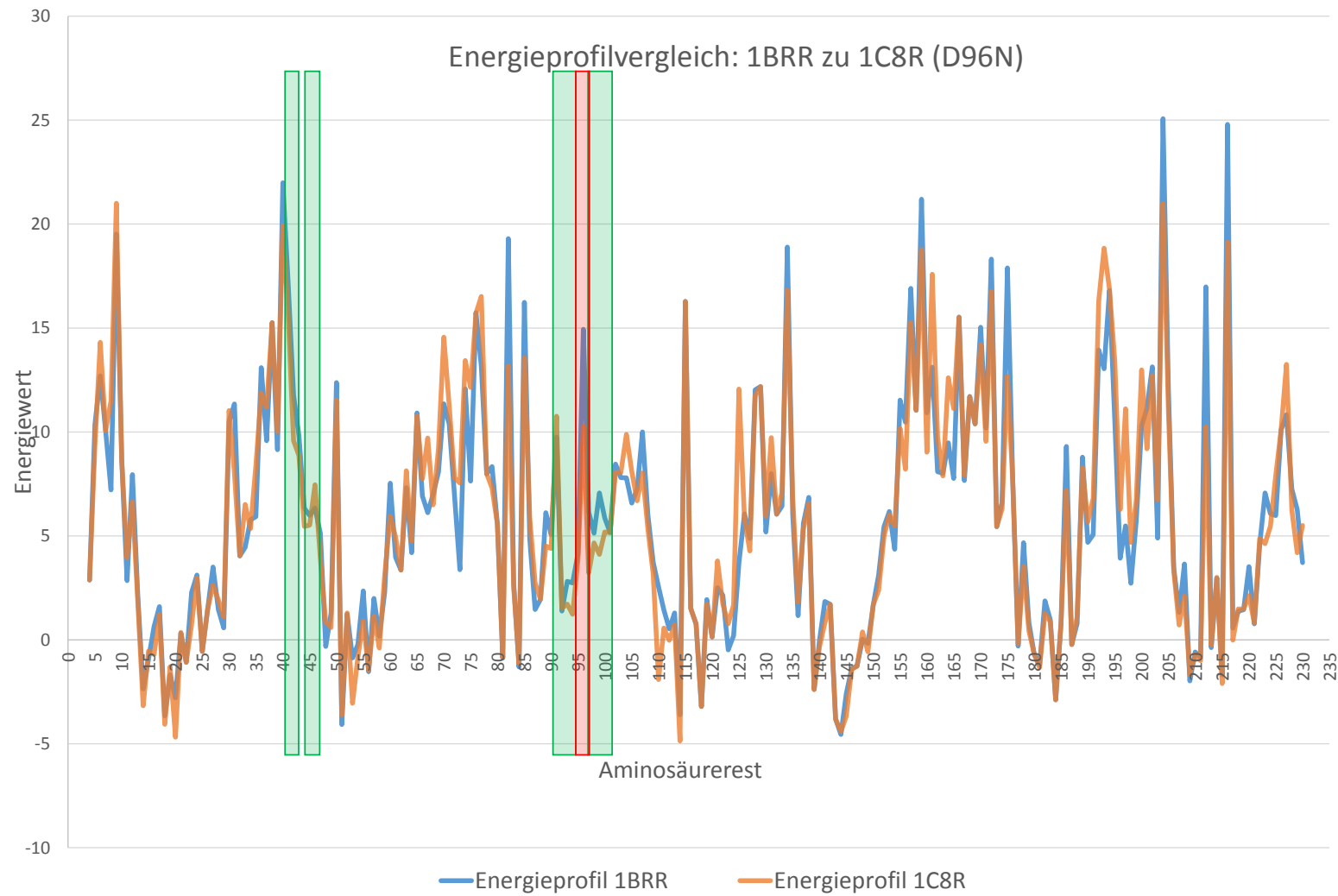


Abbildung 5.1: Vergleich des Energieprofils von Bacteriorhodopsin zum Energieprofil der Bacteriorhodopsinmutante D96N. In rot ist der Aminosäurerest der Mutation markiert, während in grün die Kontakte und somit die direkt beeinflussten Aminosäurereste markiert sind.

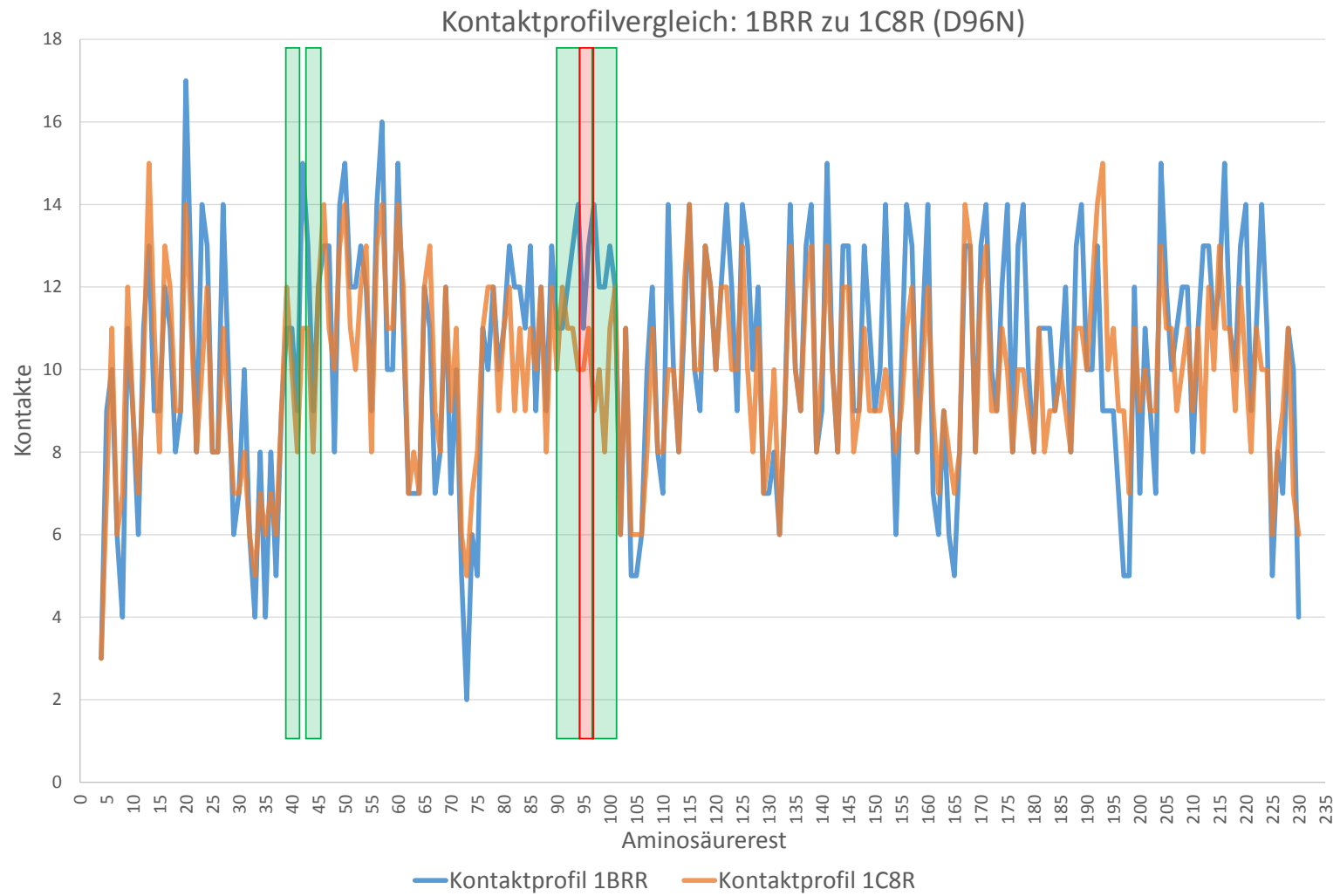


Abbildung 5.2: Vergleich des Kontaktprofils von Bacteriorhodopsin zum Energieprofil der Bacteriorhodopsinmutante D96N. In rot ist der Aminosäurerest der Mutation markiert, während in grün die Kontakte und somit die direkt beeinflussten Aminosäurereste markiert sind.

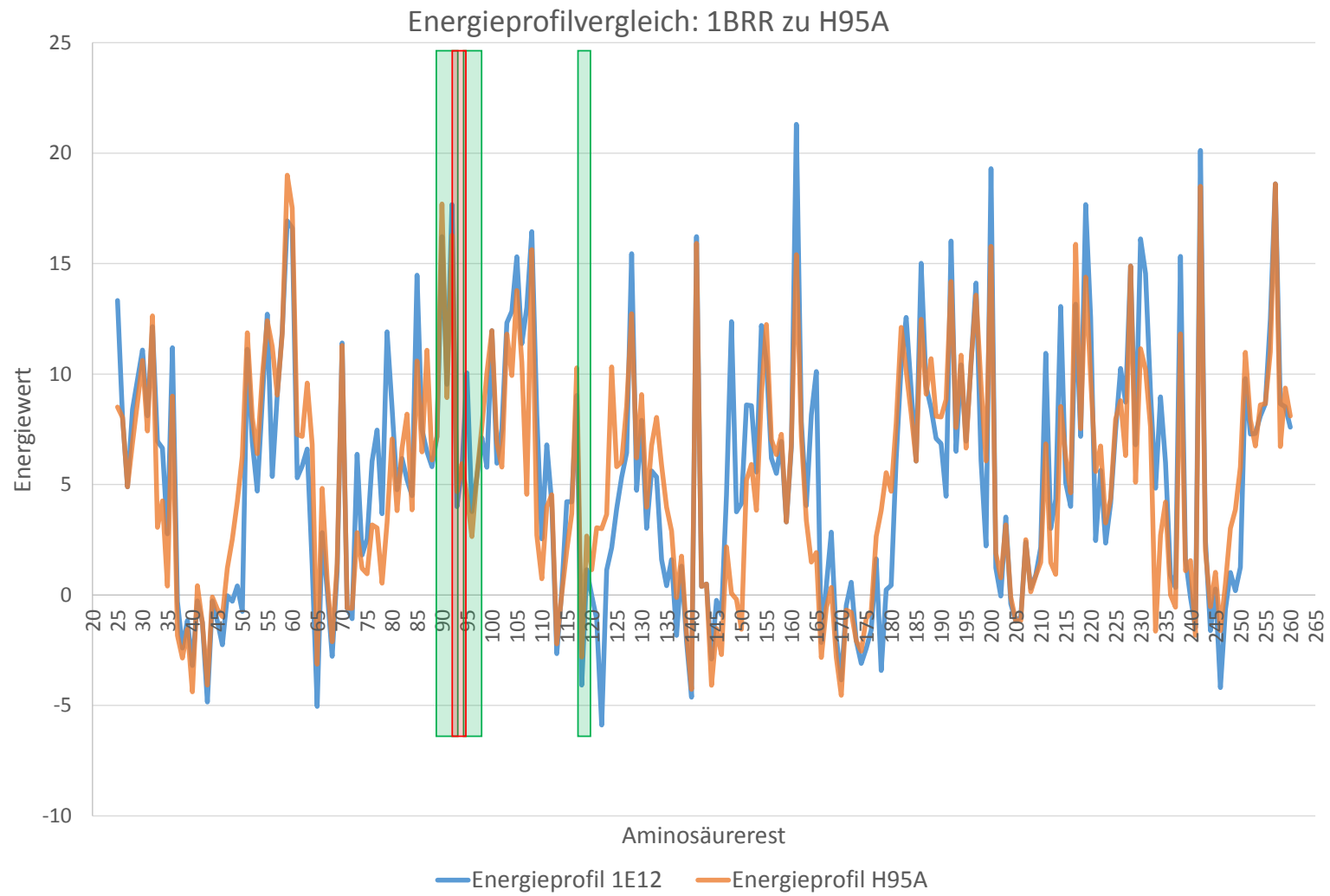


Abbildung 5.3: Vergleich des Energieprofils von Bacteriorhodopsin zum Energieprofil der Halorhodopsinmutante H95A. In rot ist der Aminosäurerest der Mutation markiert, während in grün die Kontakte und somit die direkt beeinflussten Aminosäurereste markiert sind.

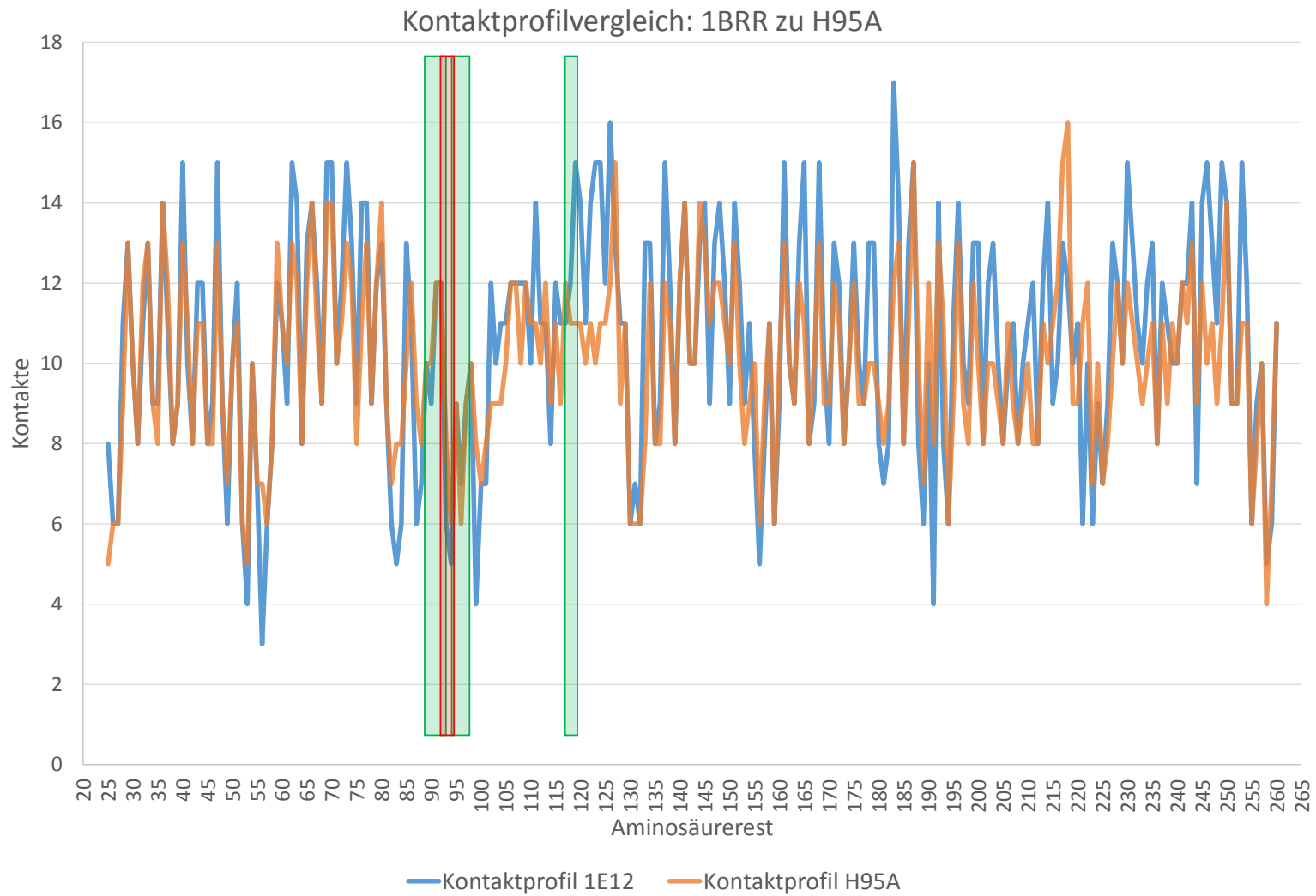


Abbildung 5.4: Vergleich des Kontaktprofils von Bacteriorhodopsin zum Energieprofil der Halorhodopsinmutante H95A. In rot ist der Aminosäurerest der Mutation markiert, während in grün die Kontakte und somit die direkt beeinflussten Aminosäurereste markiert sind.

Es lässt sich mit der hier verwendeten Vergleichsmethode keinerlei Aussage über die Auswirkungen einer veränderten Aminosäure auf die Aktivität des Proteins treffen, sondern lediglich über den Einfluss auf das Energieprofil. Hierbei muss beachtet werden, dass eine stabilisierende oder destabilisierende Mutation nicht zwangsläufig bedeutet, dass das Protein besser oder schlechter ist.

Bei der Betrachtung der Entfaltungsbarrieren fällt auf, dass es lediglich eine Verschiebung um einige Aminosäuren gibt, welche durch die Mutation entstehen können. Allerdings bleibt es hier nicht ausgeschlossen, dass auch die Veränderungen im Kontaktprofil einen starken Einfluss haben. Die Vorhersage der Entfaltungsbarrieren gibt weiterhin Einblicke in die Entstehung von Entfaltungsbarrieren. Für vier der betrachteten Mutationen von Bacteriorhodopsin gibt es experimentelle Daten, welche sich nicht von denen des nativen Proteins unterscheiden. Hier zu nennen sind die Mutationen P50A, M56A, P91A und P186A. Diese Mutationen führen zu keinerlei Veränderungen in den experimentell ermittelten Entfaltungsbarrieren, allerdings verursachen sie zum Teil stark veränderte Energieprofile. Dies lässt die Aussage zu, dass eine einzelne Mutation innerhalb sowieso außerhalb einer Energiebarriere, keinerlei Einfluss auf deren Ausbildung hat. Somit bedarf es weiterer Informationen, um eine Barriere noch genauer zu definieren und vorhersagen zu können, als die energetischen und strukturellen Eigenschaften von Proteinen.

5.2 Vergleich Bacterio- und Halorhodopsin

Ein besonderes Augenmerk muss auf dem Vergleich der beiden Proteine Bacterio- und Halorhodopsin gelegt werden, da sich diese strukturell sehr ähnlich sind. Laut FATCAT [YG04] sind sie sogar identisch, was ihr 3D-Strukturalignment anbelangt. Dieses 3D-Strukturalignment bezieht sich auf die Backbone der Proteine und deren C_{α} -Atome. Diese werden durch den Algorithmus so übereinander gelegt, dass der Abstand des C_{α} -Atoms von Protein A möglichst gering zu dem von Protein B ist. Dieser wird über alle C_{α} -Atome durchgeführt und die geringsten Abstände ermittelt. Der Umstand, dass Bacterio- und Halorhodopsin laut diesem Vergleich identisch sind, könnte durch das Runden ab einer gewissen Nachkommastelle zustande kommen. Um eine Aussage über die mögliche Vererbung und Homologie der Energiebarrieren in Proteinen treffen zu können, ist diese Ähnlichkeit allerdings essentiell. In der folgenden Abbildung sind die in Kapitel 3 aufgeführten Barrierenvorhersagen in die 3D-Struktur eingetragen und die beiden betrachteten Proteine 3D-strukturaligniert.

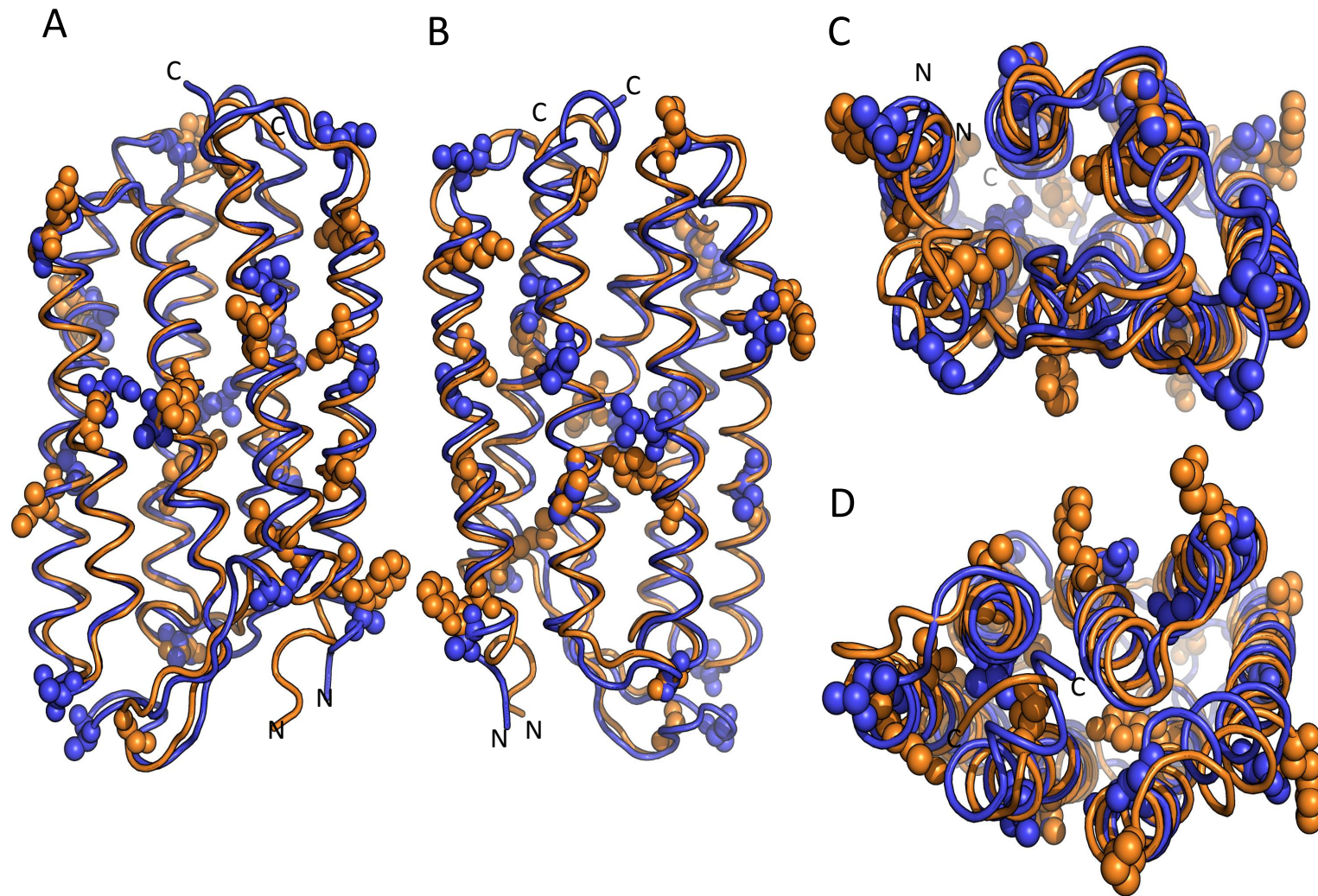


Abbildung 5.5: Vergleich der Barrierenvorhersagen anhand der Proteine Bacterio- und Halorhodopsin. Bacteriorhodopsin ist dabei in orange und Halorhodopsin in blau dargestellt. Die vorhergesagten Barrieren werden innerhalb des Proteins als Sphäre angezeigt. Abbildung **A** zeigt die Proteine von der Vorderseite, während Abbildung **B** sie von der Rückseite zeigt. Abbildung **C** zeigt die Proteine von oben und Abbildung **D** von unten. N und C zeigen jeweils die Termini in den Abbildungen an.

Es ist klar erkennbar, dass die Vorhersagen in den Helices und auch in den verbindenden Loops in ähnlichen Bereichen liegen. Zum Teil liegen die Vorhersagen, strukturell betrachtet, genau übereinander. Bei dieser Ähnlichkeit der beiden Proteine war dies zu erwarten. Interessanter ist die Tatsache, dass die falsch positiv Vorhersagen (vgl. Kapitel 3) in den selben strukturellen Bereichen der Helices der beiden Proteine liegen. Bei beiden Proteinen liegen alle falschen Vorhersagen in den Helices B,D und F. Bei Halorhodopsin wird lediglich ein falsch positiver Wert mehr vorhergesagt als bei Bacteriorhodopsin. Diese Tatsachen sprechen dafür, dass der Algorithmus zur Vorhersage von Entfaltungsbarrieren in α -helikalen Membranproteinen, funktioniert, und keine vermeintlich zufälligen Ergebnisse produziert.

5.3 Vergleich humane Aquaporine 1

Die drei unterschiedlich aufgeklärten humanen Aquaporine 1 haben nur einen kleinen Unterschied in der Auflösung wobei 3,8Å die schlechteste und 3,58Å die beste Auflösung ist. Hierbei ist allerdings auffällig, dass sich die Barrierenvorhersagen unterscheiden. Dies könnte an der unterschiedlichen Güte der Aufklärung liegen. Auch kann es sein, dass es an der Kristallisation und dem Moment der Kristallisation des Proteins liegt. Dies lässt sich zum Beispiel dadurch erklären, dass bei unterschiedlichen Aufklärungen, jeweils unterschiedliche momentane Bewegungen des Proteins auskristallisiert wurden, was letztendlich zu unterschiedlichen Ergebnissen bei der Aufklärung führte. Die folgenden zwei Abbildungen zeigen das Ausmaß der Unterschiede in den Aufklärungen anhand der Vergleiche der Energie- und Kontaktprofile.

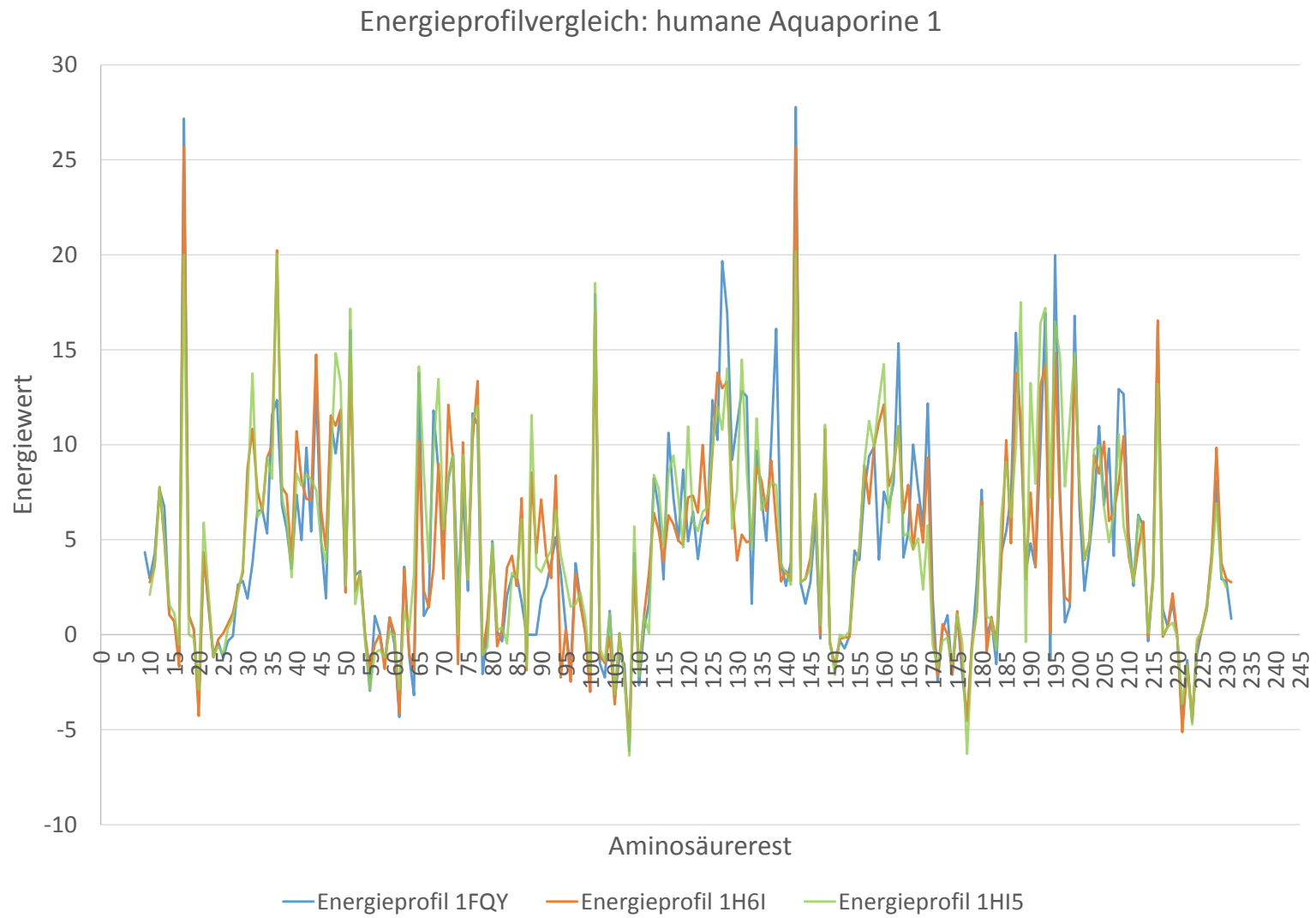


Abbildung 5.6: Grafische Darstellung der Energiewerte der einzelnen Aminosäuren der drei humanen Aquaporine 1

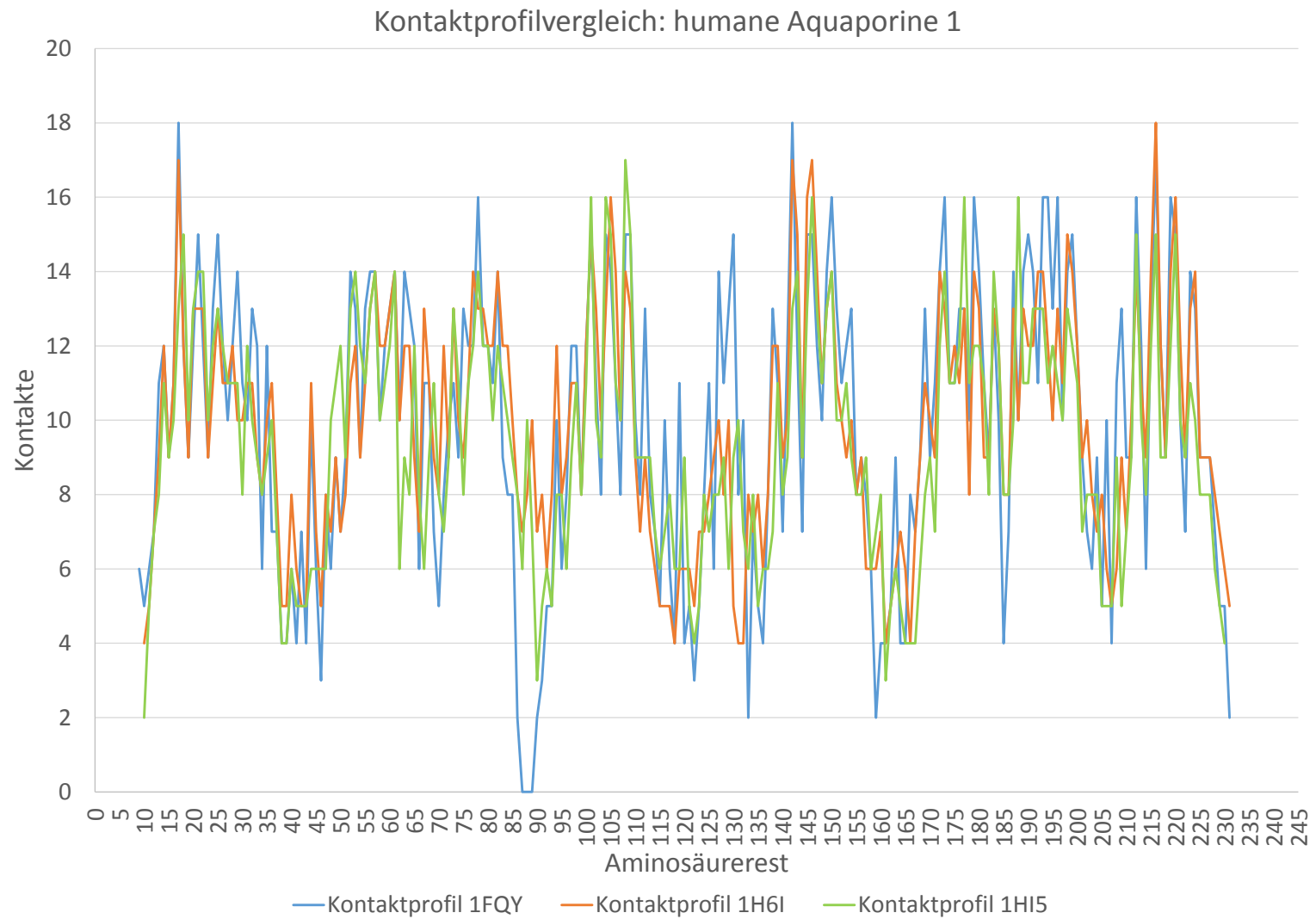


Abbildung 5.7: Grafische Darstellung der Kontaktwerte der einzelnen Aminosäuren der drei humanen Aquaporine 1

Trotz aller Unterschiede in den verwendeten Profilen sind die Differenzen in den Ergebnissen der Vorhersage nur sehr gering und unterscheiden sich nur um wenige Aminosäuren. Der größte Unterschied zwischen den Vorhersagen der einzelnen Proteine besteht darin, dass einige Vorhersagen getroffen werden oder nicht. Das Nichtvorhandensein der Vorhersagen kann man hierbei auf die Veränderungen im Kontakt- und somit auch im Energieprofil zurückführen (vgl. 2.3.1). Auch stellt sich durch diesen Vergleich die Frage, ob ein Algorithmus mit schwächeren Schwellwerten und somit mehr Vorhersagen eine Angleichung dieser unterschiedlich aufgeklärten Proteine zur Folge hat oder ob man bei Überschneidungen der erzielten Ergebnisse von gesicherten Barrieren und bei unterschiedlichen Ergebnissen von eventuellen Entfaltungsbarrieren sprechen kann.

5.4 Vergleich Aquaporine 1 bis 11

Beim Vergleich der homologen Proteine der Aquaporine 1 bis 11 war es vor allen Dingen wichtig, eine Aussage über die Vererbung der Entfaltungsbarrieren treffen zu können. Demnach soll diese Betrachtung Aufschluss über die mögliche Homologie von Barrieren geben. Einen ersten Einblick geben hier die Ergebnisse der Diskussion beim Vergleich der Proteine Halo- und Bacteriorhodopsin in Kapitel 5.2. Auffällig ist die Ähnlichkeit der Barrierenvorhersagen aller Aquaporine sowohl bei den 3D-strukturvorhergesagten als auch bei den 3D-strukturbestimmten Proteinen (vgl. Kapitel 3.3.2 und Tabellen 3.25 und 3.26). Hierbei fällt auf, dass sich die Vorhersagen nur leicht verschieben bzw. nur in wenigen Proteinen einige Vorhersagen nicht getroffen wurden. Dabei macht es keinen Unterschied, wie lang das untersuchte Protein ist. Die längeren der Aquaporine überschneiden sich bei den Vorhersagen genauso wie die kürzeren, da hier lediglich die Propeptide länger sind. Der eigentliche aktive Teil des Proteins, der als Wassertunnel fungiert, ist strukturell fast identisch. Um die strukturellen Überschneidungen der Vorhersagen aus den Tabellen 3.25 und 3.26 grafisch darzustellen, dient die folgende Abbildung 5.8.

Vergleich Humanes Aquaporin

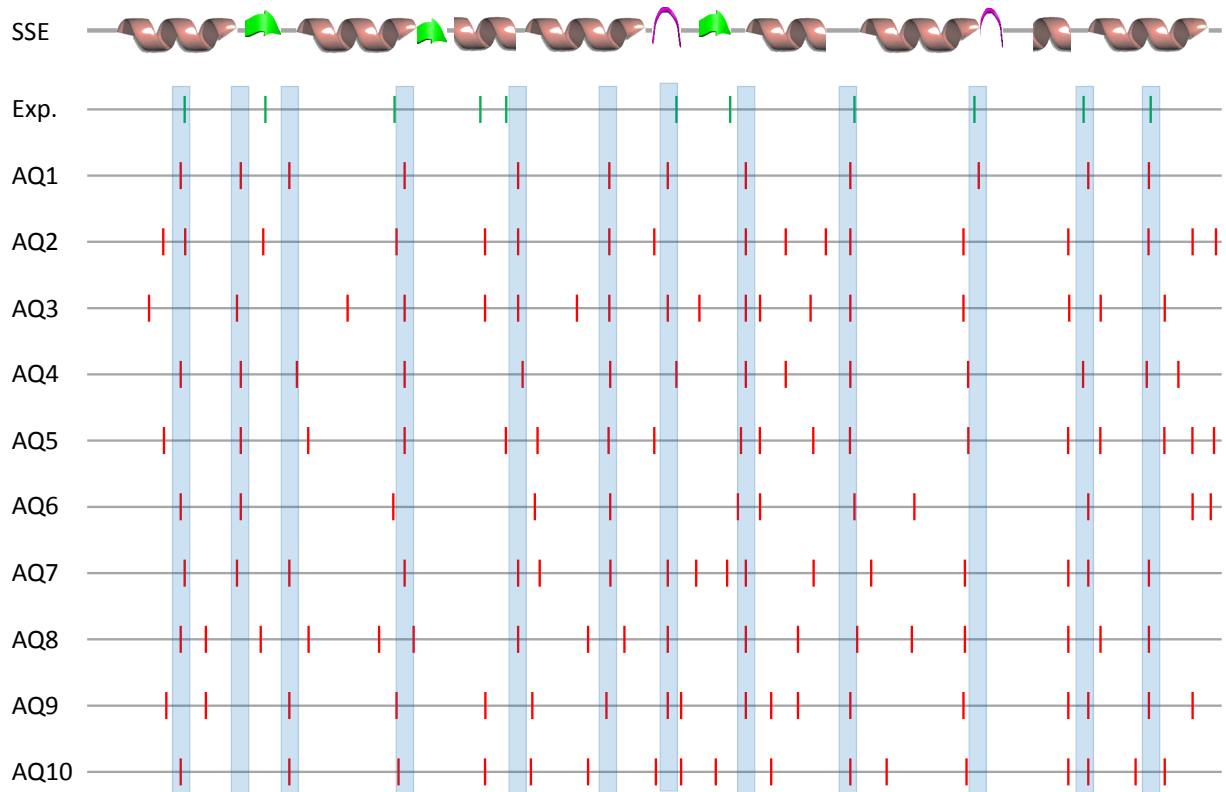


Abbildung 5.8: Grafische Darstellung des Vergleichs der unterschiedlichen Vorhersagen der humanen Aquaporine. Hierbei werden die Vorhersagen der Aquaporine (AQ) von eins bis elf schematisch gegen die Sekundärstrukturelemente (SSE) aufgetragen. Die Vorhersagen sind hierbei in rot dargestellt. Zusätzlich werden experimentelle Daten (Exp.) schematisch gegen die SSE aufgetragen. Dabei werden die experimentell ermittelten Entfaltungsbarrieren in grün dargestellt. Die blauen Balken sollen verdeutlichen, ob eine Vorhersage der anderen Aquaporine im selben Bereich liegt wie die von Aquaporin 1.

Bei der Abbildung muss beachtet werden, dass nur ein struktureller Vergleich der Vorhersagen von Entfaltungsbarrieren der Aquaporine 1 bis 10 möglich war. Dies liegt daran, dass das 3D-Strukturalignment des Aquaporin 11 zwar auf das von Aquaporin 1 passt, sich die N- bzw. C-Termini jedoch nicht übereinander befinden, sondern inmitten des anderen Proteins. Aus diesem Grund kann kein Vergleich zwischen den Sekundärstrukturelementen und den Entfaltungsbarrieren hergestellt werden.

Durch diesen strukturellen Vergleich der Vorhersagen könnte man durch die Überlagerung dieser, eine bestimmtere Aussage über die Güte und Verlässlichkeit einer Barrierenvorhersage erhalten. So ist es möglich, dass Vorhersagen, welche in allen Proteinen getroffen werden, wirklich über die Homologie erklärbar sind und dementsprechend auch wirklich in der Natur vorkommen, sozusagen eine Konsensusvorhersage bilden. Es ist allerdings nicht auszuschließen, dass durch Besonderheiten im Protein eine einzelne Vorhersage zustande kommt, welche auch experimentell nachgewiesen werden

würde, was an dieser Stelle allerdings nicht überprüfbar ist.

Wie bereits bei dem Vergleich der Proteine Bacterio- und Halorhodopsin festgestellt wurde, kann keine exakte Aussage über die Vererbung von Entfaltungsbarrieren getroffen werden. Es muss nicht unbedingt an der Homologie der Proteine liegen, dass die Übereinstimmungen der Vorhersagen der Entfaltungsbarrieren so groß sind sondern es kann auch sein, dass es strukturbedingt ist und jedes andere Protein mit ähnlicher Struktur, welches nicht homolog ist, die selben oder ähnliche Entfaltungsbarrieren aufweist.

5.5 Vergleich Vorhersage mit randomisierten Daten

Auch wenn die Güte des Vorhersagealgorithmus für Entfaltungsbarrieren in α -helikalen Membranproteinen mit einem MCC und einem F1-Maß von über 0,86 schon sehr gute Ergebnisse liefert, so ist erkennbar, dass die Güte von zufälligen Ergebnissen mit randomisierten Daten von ca. 0,69, nur geringfügig schlechter ist. Dies wirkt weniger überraschend, wenn man bedenkt, dass die Vorhersagen, wie bereits in 4.3 erwähnt wurde, nicht komplett auf Zufall beruhen, sondern auf Wahrscheinlichkeiten, die auf den Ergebnissen des gut funktionierenden Vorhersagealgorithmus basieren. Des Weiteren ist der Parameter, welcher für die Güte verantwortlich ist (vgl. Kapitel 2.2) sehr weitläufig gefasst und führt somit auch bei einem zufälligen Ergebnis häufig zu richtigen Vorhersagen. Auch liegen die experimentell ermittelten Barrieren häufig sehr nah beieinander, was dazu führt, dass ein zufällig ermittelter Wert teilweise zwei unterschiedlich experimentell ermittelte Entfaltungsbarrieren erklärt und somit zu einer hohen Güte führt. Der größte Unterschied zwischen beiden Verfahren werden die Vorhersagen innerhalb der Abschnitte des Proteins sein, in denen keine experimentellen Daten und auch die falsch positiven Vorhersagen liegen. Allerdings werden in diesen Bereichen bei dem zufälligen Verfahren hier mehr falsche Vorhersagen liegen als bei der ursprünglichen Vorhersage. Diese Aussage unterstreicht die Tatsache, dass eine komplett auf Zufall beruhende Vorhersage eine nur um ca. 0,04 schlechtere Güte hat, als die Vorhersage, welche auf den erzeugten Wahrscheinlichkeiten beruht. Wenn man die Aussage auf die Standardabweichung der Gütemaße bezieht, so erkennt man, dass die zufällige Vorhersage signifikant schlechter ist, als die auf berechneten Wahrscheinlichkeiten beruhende. Im Verhältnis zu der vollkommen zufälligen Vorhersage ist der Vorhersagealgorithmus ausgehend von einem MCC von 0,868 und einem F1-Maß von 0.869 für den nativen Zustand der Proteine 13,1 mal besser bzw. 12 mal besser als die zufällige Vorhersage (vgl. 4.3). Somit kann, bei einem Unterschied von ca. 13 Standardabweichungen, davon ausgegangen werden, dass die Vorhersage der Entfaltungsbarrieren kein Zufall sind, und der Vorhersagealgorithmus eine verlässliche Güte besitzt.

6 Molekularbiologische Interpretation

Die auftretenden Unterschiede zwischen den aufgeklärten Proteinen, sowohl bei Bacteriorhodopsin, Halorhodopsin und Aquaporin 1 als auch bei deren Mutationen, könnten durch unterschiedliche Kristallisation oder durch natürliche Bewegungen des Proteins zustande kommen.

Durch den Vergleich von Bacterio- und Halorhodopsin und durch den Vergleich der Aquaporine 1 bis 11, kann man sagen, dass sich Entfaltungsbarrieren in einer homologen Reihe, aufgrund der Strukturähnlichkeiten, nicht sehr groß unterscheiden. Allerdings ergibt sich an dieser Stelle die Frage, ob es an der Homologie der Proteine liegt, die zu einer Ähnlichkeit in den Entfaltungsbarrieren führt oder ob die Struktur die Entfaltungsbarrieren bedingt und somit auch andere, nicht verwandte Proteine ähnliche oder sogar die gleichen Entfaltungsbarrieren aufweisen. Dagegen spricht allerdings die Tatsache, dass die Entfaltungsbarrieren von Bacterio- oder Halorhodopsin und Aquaporin 1 nicht vergleichbar sind, obwohl sich die Strukturen sehr ähneln.

Im Punkt 5.2 wurde der Vergleich zwischen Halo- und Bacteriorhodopsin angestellt, um die Verlässlichkeit des Vorhersagealgorithmus zu bestätigen. Da sich die falsch positiven Vorhersagen auffällig oft in den Helices befanden, lag eine Überprüfung des Experiments beider Proteine nahe. Bei dieser Überprüfung stellte sich heraus, dass es die Möglichkeit der spontanen Entfaltung innerhalb der Membran gibt und auch eine Temperatur- und Zuggeschwindigkeitsabhängigkeit [JSH⁺04] vorhanden ist. Weiterhin ist auffällig, dass sich alle falsch positiven Vorhersagen bei Halo- und Bacteriorhodopsin innerhalb bzw. am Anfang und Ende von Helices befinden, welche komplett ohne experimentelle Barrieren vorliegen. Auch sind die Bereiche ohne experimentellen Daten für ein Protein von ca. 240 Aminosäuren sehr lang (ca. 40 bis 50 Aminosäuren). Des Weiteren sind die Kräfte, die benötigt werden, um die vor diesen Bereichen liegende Barrieren zu lösen, mit teilweise 200 pN sehr hoch. Teilweise wird mit fast 3-facher Kraft gezogen bis die Barriere gelöst ist. Es könnte sein, dass die Kraft, welche vor den Helices mit den falsch positiven Vorhersagen wirkt, dazu führt, dass eine spontane Selbstentfaltung der nachfolgenden Helix stattfindet. Auch könnte es sein, dass sich Barrieren innerhalb der Helix entfalten, da die Energie gering ist und lediglich eine schwache Barriere ausbilden. Die übertragene Energie der Entfaltung vorher könnte somit ausreichen, um die Barriere vor ihrer Detektion aufzulösen. Somit ist die Hypothese legitim, dass der Vorhersagealgorithmus sogar noch besser arbeitet als das Experiment auf den ersten Blick vermuten lässt, da er sogar im Experiment nicht detektierbare Entfaltungsbarrieren erkennt und vorhersagt und somit sogar eine Verbesserung gegenüber des Experiments darstellt. Diese Hypothese lässt sich zum jetzigen Zeitpunkt allerdings nicht überprüfen.

Weiterhin muss festgehalten werden, dass im Falle des Vorhersagealgorithmus das Experiment die entscheidende Variable über die Güte des Algorithmus ist. Wie in den bisherigen Abschnitten bereits erwähnt, spielt es eine entscheidende Rolle, welche Auf-

klärung eines Proteins verwendet wird und welche Auflösung diese besitzt. Zwar ist die Güte in einem gewissen Maße proteinübergreifend, wie man in der Tabelle des Vergleiches der Aquaporine 1 entnehmen kann, dennoch spielt das verwendete Protein eine große Rolle bei der Wahl der Parameter. Eine Möglichkeit, wie man diesem Umstand eventuell aus dem Weg gehen könnte, wäre die Aufklärung von Membranproteinen nicht über Röntgenstrukturanalyse sondern über die Kernspinresonanzspektroskopie. Dies hätte den Vorteil, dass in den verwendbaren Daten direkt eine Bewegung des Proteins enthalten ist, was eine Vorhersage spezifischer und, über die momentan verschiedenen Aufklärungen hinweg, besser machen würde.

Mit den in dieser Arbeit gesammelten Ergebnissen und Erkenntnissen, lässt sich die Hypothese aufstellen, dass ein Membranprotein während der Entfaltung durch eine externe Kraft keiner lateralen Verschiebung in der Membran ausgesetzt ist und während des Experiments in der Position vom Beginn des Ziehens an verharrt. Ein Beleg, der diese Aussage unterstreicht, ist die Tatsache, dass die Betrachtung des nativen Zustands eine höhere Güte erzielt als bei dem Versuch, das Experiment zu simulieren. Für diese These spricht außerdem die Tatsache, dass Bacteriorhodopsin, wenn das Herausreißen aus der Membran nach der fünften Helix, abgebrochen wird, eine recht lange Zeit benötigt, um sich zurück in den nativen Zustand zu falten [JKO⁺03]. Hierbei scheint sich das Protein vor der spontanen Faltung, in einer Starre zu befinden, die es erst zu überwinden gilt.

7 Zusammenfassung

Membranproteine sind wichtige Bestandteile eines jeden Organismus und verantwortlich für viele Steuer- und Regulationsmechanismen der Zellen. So groß auch das Interesse an ihnen ist, so schwierig sind sie in der Handhabung. Durch ihre speziellen Eigenschaften und Bedürfnisse, was die Umgebung angeht, sind experimentelle Methoden, die auf globuläre Proteine ausgelegt sind, nur sehr bedingt auf Membranproteine zu übertragen. Dies führt dazu, dass es nur wenige und meist zeit- und kostenintensive Methoden gibt, welche einen direkten Erkenntnisgewinn aus Membranproteinen gewinnen.

In dieser Arbeit ging es um die Entwicklung und Anwendung eines Vorhersagealgorithmus für Entfaltungsbarrieren in α -helikale Membranproteinen. Ziel war es dabei, eine kosten- und zeitgünstige Alternative zum Experiment zu entwickeln, welche mit einer verlässlichen Güte Entfaltungsbarrieren innerhalb von α -helikalen Membranproteinen vorhersagt. Des Weiteren war es, sowohl Ziel neue Erkenntnisse über die Vererbung und somit Homologie von Entfaltungsbarrieren zu erlangen als auch die Auswirkungen von Mutationen auf die Ausbildung von Entfaltungsbarrieren zu erlangen.

Zusammenfassend kann man sagen, dass diese Arbeit einige Erkenntnisse über Membranproteine generieren konnte. Der Vorhersagealgorithmus konnte soweit etabliert werden, dass die Güte, im Vergleich zu einem zufälligen Experiment, viel höher und genauer ist. Allerdings ist die Güte dabei nur so gut wie das Experiment und wird auch durch dieses bestimmt. Des Weiteren kann man sagen, dass verschieden aufgeklärte Proteine dazu führen, dass unterschiedliche Vorhersagen erzeugt werden, obwohl es sich um das gleiche Protein handelt. Hierbei ist allerdings zu erkennen, dass es nur geringfügige Schwankungen gibt, welche eventuell sogar durch die natürlichen Bewegungen des Proteins erklärt werden können, bei der Kristallisation allerdings zu unterschiedlichen Ergebnissen der 3D-Struktur führen.

Die Arbeit hat weiterhin Einblicke in einige Punktmutationen gewährt und konnte eine klare Aussage über die Auswirkungen dieser Mutationen auf die Ausbildung von Entfaltungsbarrieren treffen. Auch wurde ein globaler Einfluss, unter Zuhilfenahme des Gesamtpotentials der Energieprofile, ermittelt. Da Aufgrund von unterschiedlichen Aufklärungen von Proteinen oder der 3D-Strukturvorhersage Schwankungen in die Betrachtung einwirken, welche nicht zwangsläufig auf die Mutation zurückzuführen sein müssen, ist die Aussage allerdings nicht so konkret zu behandeln, wie die Aussage über das Ausbilden von Entfaltungsbarrieren.

Durch die Betrachtung verschiedener homologer Proteine der Aquaporine konnte eine Aussage über die mögliche Vererbung und somit Homologie von Entfaltungsbarrieren getroffen werden. Die betrachteten Proteine lassen keinen anderen Schluss zu, als den, dass obwohl die Entfaltungsbarrieren zwar auf eine gewisse Art individuell für jedes Protein sind, dennoch die Überschneidungen groß genug sind, um eine Konsensusvorhersage über alle Proteine treffen zu können. Dies führt letztendlich zu der Vermutung,

das Entfaltungsbarrieren über die Homologie erklärbar und vererbbar sind. Um diese Aussage letztendlich verifizieren zu können, muss allerdings ausgeschlossen werden, dass die Ähnlichkeit der vorhergesagten Entfaltungsbarrieren nicht strukturell bedingt werden, sondern wirklich ausschließlich aufgrund der Homologie.

8 Ausblick

Auch wenn die vorliegende Arbeit einen sehr guten Einblick in die Entfaltungsbarrieren von α -helikalen Membranproteinen gibt, so muss an einigen Stellen weiter geforscht werden, um die Aussagen zu präzisieren, oder um eine neue Aussage treffen zu können.

Im Anschluss an diese Arbeit ist es notwendig, auch andere, nicht homologen Rhodopsine zu betrachten. Dies dient der Verifizierung der Aussage, dass die Entfaltungsbarrieren über die Homologie hinweg vererbt werden und nicht von der Struktur abhängig sind. Als kleiner Einblick in dieses Vorgehen dient die folgende Abbildung, welche verdeutlichen soll, dass die Homologie von Proteinen einen direkten Einfluss auf die Vorhersage von Entfaltungsbarrieren und somit auch auf ihre Ausbildung in der Natur hat. Die Abbildung zeigt zum einen den Vergleich von Archæorhodopsin 1 und 2 [EYM⁺06] zu Bacteriorhodopsin und zum anderen den Vergleich von Anabaena sensory rhodopsin [VST⁺04] zu Bacteriorhodopsin.

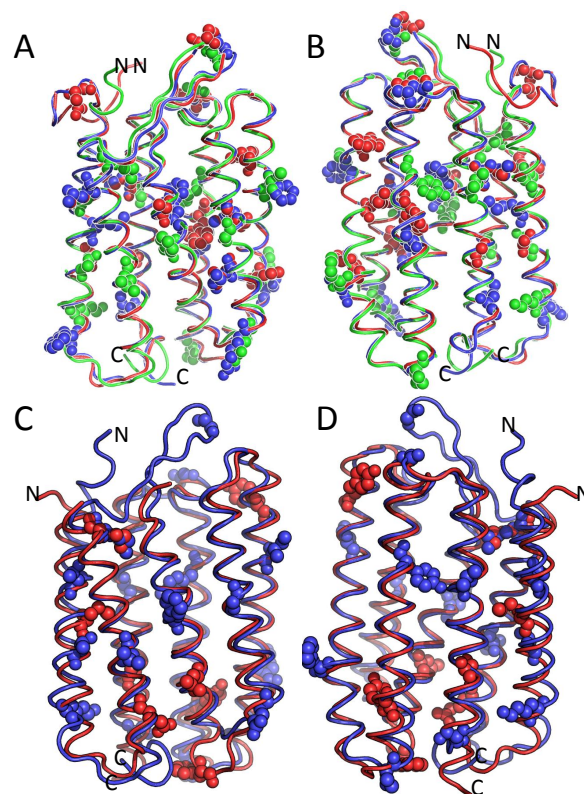


Abbildung 8.1: Vergleich der Vorhersagen in verschiedenen Rhodopsinen. Dabei zeigt **A** den Vergleich der Archæorhodopsin 1 (rot) und 2 (blau) zu Bacteriorhodopsin (grün) von vorne, während **B** den Vergleich von hinten zeigt. **C** zeigt den Vergleich von Anabaena sensory rhodopsin (blau) zu Bacteriorhodopsin (rot) von vorne und **D** von hinten. N und C bezeichnen hierbei die entsprechenden Termini der Proteine.

Zur besseren Darstellung der Übereinstimmungen und Diskrepanzen in den Vorhersa-

gen, dient die Auftragung der Vorhersagen gegen die Sekundärstruktur der Proteine in Abbildung 8.2.

Vorhergesagte Entfaltungsbarrieren: Vergleich Archaerhodopsin 1, 2 und Anabnea Sensory Rhodopsin zu Bacteriorhodopsin

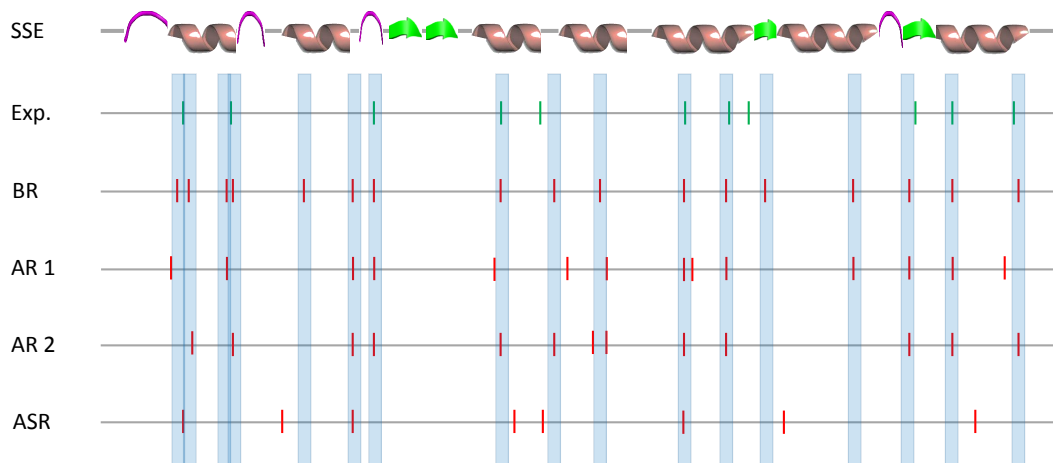


Abbildung 8.2: Vergleich der vorhergesagten Entfaltungsbarrieren der Proteine Bacteriorhodopsin (BR), Archaerhodopsin 1 (AR 1), Archaerhodopsin 2 (AR 2) und Anabaena Sensory Rhodopsin (ASR). Dabei werden die Vorhersagen gegen die Sekundärstrukturelemente (SSE) aufgetragen. Zum Vergleich sind auch die experimentell ermittelten Barrieren (Exp.) aufgetragen. Die blauen Balken dienen zur Orientierung der Überschneidungen der Vorhersagen im Vergleich zu den Vorhersagen von Bacteriorhodopsin.

Man erkennt, dass die homologen Archaerhodopsine 1 und 2 im Vergleich zu Bacteriorhodopsin fast identische Vorhersagen besitzen. Die Vorhersagen des Anabaena sensory rhodopsin unterscheiden sich grundlegend von denen des Bacteriorhodopsins. Damit liegt die Vermutung nahe, dass Entfaltungsbarrieren tatsächlich über die Homologie vererbt werden und die Struktur nur bedingt etwas mit der Übereinstimmung zu tun hat. Um diese Aussage noch weiter zu untermauern, sollte nach der Betrachtung der unterschiedlichsten Rhodopsine auch eine Untersuchung von anderen Proteinen mit sieben transmembranen Helices folgen, sogenannten 7TM-Rezeptoren.

Ein weiterer Untersuchungsaspekt könnte die Primärstruktur der Proteine betreffen. Für einzelne Aminosäurereste ist schon lange eine Konservierung über die Homologie hinweg bekannt. Die Frage ist, ob eine solche Konservierung von Entfaltungsbarrieren auf energetischer oder struktureller Ebene stattfindet und ob regelmäßige Muster (Pattern) an der Ausbildung von Entfaltungsbarrieren beteiligt sind. Falls es solche Pattern gibt, wäre ein Untersuchungsaspekt sowohl die Zusammensetzung dieser Pattern als auch

ihre Länge, was einen direkten Rückschluss auf die Größe der Entfaltungsbarriere geben könnte. Außerdem könnte überprüft werden, ob es gewisse Aminosäuren gibt, welche auffällig häufig oder sogar immer, an der Bildung von Energiebarrieren beteiligt sind. Auch eine Verbindung zwischen konservierten Aminosäureresten und Entfaltungsbarrieren sollte in diesem Zusammenhang untersucht werden.

Um den globalen Einfluss von Mutationen in Membranproteinen besser einschätzen zu können, wäre eine Betrachtung von mehreren Mutationen im selben Protein sinnvoll. Hier könnte eine weiterführende Einschätzung der Auswirkungen auf die Stabilität des Proteins von Mutationen in Membranproteinen das Resultat sein.

Außerdem ist es notwendig den in dieser Arbeit entwickelten und verwendeten Algorithmus weiter zu verbessern. Um dies zu erreichen, könnte das momentan sehr einfach betrachtete System, um einige Komponenten erweitert werden. Eine Erweiterung könnte die nochmalige Einbringung von ANM oder Data-Mining-Tools sein, um die Beschaffenheit von Energiebarrieren besser zu verstehen. Eine weitere Möglichkeit, den Vorhersagealgorithmus zu verbessern, wäre die Untersuchung der im Experiment angewendeten Kraft und die anschließende Übertragung auf das Gesamtsystem.

Letztlich ist auch eine Verbesserung des verwendeten Energieprofils anzustreben. Das momentan verwendete Profil gibt zwar einen guten Einblick in die Energien und somit auch in die Stabilität des Proteins, stützt sich allerdings lediglich auf die Lösungsenergien der Aminosäurereste (vgl. 1.3.2). Die Verbesserung des Energieprofils könnte in einer besseren Verschränkung von Energie- und Kontaktpprofil bestehen. Das bedeutet, dass die Kontaktpräferenzen von Aminosäureresten und die tatsächlichen Kontakte eines Restes mit in die Berechnung des Energiewertes eines Aminosäurerestes und somit in das Energieprofil eingehen.

Literaturverzeichnis

- [ABKS06] ARNOLD, Konstantin ; BORDOLI, Lorenza ; KOPP, Jürgen ; SCHWEDE, Torsten: The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. In: *Bioinformatics* 22 (2006), Jan, Nr. 2, 195–201. <http://dx.doi.org/10.1093/bioinformatics/bti770>. – DOI 10.1093/bioinformatics/bti770
- [ADJ⁺01] ATILGAN, A. R. ; DURELL, S. R. ; JERNIGAN, R. L. ; DEMIREL, M. C. ; KESKIN, O. ; BAHAR, I.: Anisotropy of fluctuation dynamics of proteins with an elastic network model. In: *Biophys J* 80 (2001), Jan, Nr. 1, 505–515. [http://dx.doi.org/10.1016/S0006-3495\(01\)76033-X](http://dx.doi.org/10.1016/S0006-3495(01)76033-X). – DOI 10.1016/S0006-3495(01)76033-X
- [Buc02] BUCHNER, J.: Introduction: the cellular protein folding machinery. In: *Cell Mol Life Sci* 59 (2002), Oct, Nr. 10, S. 1587–1588
- [COM05] CISNEROS, David A. ; OESTERHELT, Dieter ; MÜLLER, Daniel J.: Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. In: *Structure* 13 (2005), Feb, Nr. 2, 235–242. <http://dx.doi.org/10.1016/j.str.2004.12.005>. – DOI 10.1016/j.str.2004.12.005
- [DAB00] DORUKER, P. ; ATILGAN, A. R. ; BAHAR, I.: Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. In: *Proteins* 40 (2000), Aug, Nr. 3, S. 512–524
- [dEG01] DE GROOT, B. L. ; ENGEL, A. ; GRUBMÜLLER, H.: A refined structure of human aquaporin-1. In: *FEBS Lett* 504 (2001), Aug, Nr. 3, S. 206–211
- [Dre08] DRESSEL, Frank: *Sequenz, Energie, Struktur - Untersuchungen zur Beziehung zwischen Primär- und Tertiärstruktur in globulären und Membran-Proteinen*, Technische Universität Dresden, Diss., 2008
- [ESLO98] ESSEN, L. ; SIEGERT, R. ; LEHMANN, W. D. ; OESTERHELT, D.: Lipid patches in membrane protein oligomers: crystal structure of the bacteriorhodopsin-lipid complex. In: *Proc Natl Acad Sci U S A* 95 (1998), Sep, Nr. 20, S. 11673–11678
- [EWMR⁺06] ESWAR, Narayanan ; WEBB, Ben ; MARTI-RENOM, Marc A. ; MADHUSUDHAN, M. S. ; ERAMIAN, David ; SHEN, Min-Yi ; PIEPER, Ursula ; SALI, Andrej: Comparative protein structure modeling using Modeller. In: *Curr Protoc Bioinformatics* Chapter 5 (2006), Oct, Unit 5.6. <http://dx.doi.org/10.1002/0471250953.bi0506s15>. – DOI 10.1002/0471250953.bi0506s15
- [EYM⁺06] ENAMI, Nobuo ; YOSHIMURA, Keiko ; MURAKAMI, Midori ; OKUMURA, Hideo ; IHARA, Kunio ; KOUYAMA, Tsutomu: Crystal structures of archaeorhodopsin-1 and -2: Common structural motif in archaeal light-driven proton pumps. In: *J Mol Biol* 358 (2006), May, Nr. 3, 675–685. <http://dx.doi.org/10.1016/j.jmb.2006.02.032>. – DOI 10.1016/j.jmb.2006.02.032
- [FYB⁺04] FAHAM, Salem ; YANG, Duan ; BARE, Emiko ; YOHANNAN, Sarah ; WHITELEGGE, Julian P. ;

- BOWIE, James U.: Side-chain contributions to membrane protein structure and stability. In: *J Mol Biol* 335 (2004), Jan, Nr. 1, S. 297–305
- [HAD13] HEINKE, Florian ; ANDRÉ, Marquardt ; DIRK, Labudde: Predicting Unfolding Barriers in α -helical Membrane Proteins using coarse-grained Representations of Protein Structures, Dynamics and Single-Molecule Force Spectroscopy. In: *CRTD Dresden*, 2013
- [HNF⁺08] HORSEFIELD, Rob ; NORDÉN, Kristina ; FELLERT, Maria ; BACKMARK, Anna ; TÖRNROTH-HORSEFIELD, Susanna ; TERWISSCHA VAN SCHELTINGA, Anke C. ; KVASSMAN, Jan ; KJELLBOM, Per ; JOHANSON, Urban ; NEUTZE, Richard: High-resolution x-ray structure of human aquaporin 5. In: *Proc Natl Acad Sci U S A* 105 (2008), Sep, Nr. 36, 13327–13332. <http://dx.doi.org/10.1073/pnas.0801466105>. – DOI 10.1073/pnas.0801466105
- [HYS⁺09] HO, Joseph D. ; YEH, Ronald ; SANDSTROM, Andrew ; CHORNY, Ilya ; HARRIES, William E C. ; ROBBINS, Rebecca A. ; MIERCKE, Larry J W. ; STROUD, Robert M.: Crystal structure of human aquaporin 4 at 1.8 Å and its mechanism of conductance. In: *Proc Natl Acad Sci U S A* 106 (2009), May, Nr. 18, 7437–7442. <http://dx.doi.org/10.1073/pnas.0902725106>. – DOI 10.1073/pnas.0902725106
- [JKO⁺03] JANOVJAK, Harald ; KESSLER, Max ; OESTERHELT, Dieter ; GAUB, Hermann ; MÜLLER, Daniel J.: Unfolding pathways of native bacteriorhodopsin depend on temperature. In: *EMBO J* 22 (2003), Oct, Nr. 19, 5220–5229. <http://dx.doi.org/10.1093/emboj/cdg509>. – DOI 10.1093/emboj/cdg509
- [JSH⁺04] JANOVJAK, Harald ; STRUCKMEIER, Jens ; HUBAIN, Maurice ; KEDROV, Alexej ; KESSLER, Max ; MÜLLER, Daniel J.: Probing the energy landscape of the membrane protein bacteriorhodopsin. In: *Structure* 12 (2004), May, Nr. 5, 871–879. <http://dx.doi.org/10.1016/j.str.2004.03.016>. – DOI 10.1016/j.str.2004.03.016
- [KAK⁺09] KIEFER, Florian ; ARNOLD, Konstantin ; KÜNZLI, Michael ; BORDOLI, Lorenza ; SCHWEDE, Torsten: The SWISS-MODEL Repository and associated resources. In: *Nucleic Acids Res* 37 (2009), Jan, Nr. Database issue, D387–D392. <http://dx.doi.org/10.1093/nar/gkn750>. – DOI 10.1093/nar/gkn750
- [KBEO00] KOLBE, M. ; BESIR, H. ; ESSEN, L. O. ; OESTERHELT, D.: Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. In: *Science* 288 (2000), May, Nr. 5470, S. 1390–1396
- [KCMP10] KAWAMURA, Shiho ; COLOZO, Alejandro T. ; MÜLLER, Daniel J. ; PARK, Paul S-H.: Conservation of molecular interactions stabilizing bovine and mouse rhodopsin. In: *Biochemistry* 49 (2010), Dec, Nr. 49, 10412–10420. <http://dx.doi.org/10.1021/bi101345x>. – DOI 10.1021/bi101345x
- [KJSM07] KEDROV, Alexej ; JANOVJAK, Harald ; SAPRA, K T. ; MÜLLER, Daniel J.: Deciphering molecular interactions of native membrane proteins by single-molecule force spectroscopy. In: *Annu Rev Biophys Biomol Struct* 36 (2007), 233–260. <http://dx.doi.org/10.1146/annurev.biophys.36.040306.132640>. – DOI 10.1146/annurev.biophys.36.040306.132640
- [KP98] KOHAVI, R. ; PROVOST, F.: Glossary of terms. In: *Journal of Machine Learning* 30 (1998), S. 271–274

- [KST13] KOZMA, Dániel ; SIMON, István ; TUSNÁDY, Gábor E.: PDBTM: Protein Data Bank of trans-membrane proteins after 8 years. In: *Nucleic Acids Res* 41 (2013), Jan, Nr. D1, D524–D529. <http://dx.doi.org/10.1093/nar/gks1169>. – DOI 10.1093/nar/gks1169
- [LSR⁺99] LUECKE, H. ; SCHOBERT, B. ; RICHTER, H. T. ; CARTAILLER, J. P. ; LANYI, J. K.: Structural changes in bacteriorhodopsin during ion transport at 2 angstrom resolution. In: *Science* 286 (1999), Oct, Nr. 5438, S. 255–261
- [MAE97] MÜLLER, D. J. ; AMREIN, M. ; ENGEL, A.: Adsorption of biological molecules to a solid support for scanning probe microscopy. In: *J Struct Biol* 119 (1997), Jul, Nr. 2, 172–188. <http://dx.doi.org/10.1006/jsbi.1997.3875>. – DOI 10.1006/jsbi.1997.3875
- [MFS⁺03] MÖLLER, Clemens ; FOTIADIS, Dimitrios ; SUDA, Kitaru ; ENGEL, Andreas ; KESSLER, Max ; MÜLLER, Daniel J.: Determining molecular forces that stabilize human aquaporin-1. In: *J Struct Biol* 142 (2003), Jun, Nr. 3, S. 369–378
- [MIL53] MILLER, S. L.: A production of amino acids under possible primitive earth conditions. In: *Science* 117 (1953), May, Nr. 3046, S. 528–529
- [MKO⁺02] MÜLLER, Daniel J. ; KESSLER, Max ; OESTERHELT, Filipp ; MÖLLER, Clemens ; OESTERHELT, Dieter ; GAUB, Hermann: Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. In: *Biophys J* 83 (2002), Dec, Nr. 6, 3578–3588. [http://dx.doi.org/10.1016/S0006-3495\(02\)75358-7](http://dx.doi.org/10.1016/S0006-3495(02)75358-7). – DOI 10.1016/S0006-3495(02)75358-7
- [MMH⁺00] MURATA, K. ; MITSUOKA, K. ; HIRAI, T. ; WALZ, T. ; AGRE, P. ; HEYMAN, J. B. ; ENGEL, A. ; FUJIYOSHI, Y.: Structural determinants of water permeation through aquaporin-1. In: *Nature* 407 (2000), Oct, Nr. 6804, 599–605. <http://dx.doi.org/10.1038/35036519>. – DOI 10.1038/35036519
- [NH97a] NAUMER, Hans ; HELLER, Wolfgang: *Untersuchungsmethoden in der Chemie Einführung in die moderne Analytik* ISBN: 3-13-681403-7. Bd. 3. erweiterte und neubearbeitete Auflage. Georg Thieme Verlag Stuttgart - New York, 1997. – S. 393–403
- [NH97b] NAUMER, Hans ; HELLER, Wolfgang: *Untersuchungsmethoden in der Chemie Einführung in die moderne Analytik* ISBN: 3-13-681403-7. Bd. 3. erweiterte und neubearbeitete Auflage. Georg Thieme Verlag Stuttgart - New York, 1997. – S. 166–187
- [OS74] OESTERHELT, D. ; STOECKENIUS, W.: Isolation of the cell membrane of Halobacterium halobium and its fractionation into red and purple membrane. In: *Methods Enzymol* 31 (1974), S. 667–678
- [OSB⁺04] OKADA, Tetsuji ; SUGIHARA, Minoru ; BONDAR, Ana-Nicoleta ; ELSTNER, Marcus ; ENTEL, Peter ; BUSS, Volker: The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. In: *J Mol Biol* 342 (2004), Sep, Nr. 2, 571–583. <http://dx.doi.org/10.1016/j.jmb.2004.07.044>. – DOI 10.1016/j.jmb.2004.07.044
- [Pow11] POWERS, David M W.: Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. In: *Journal of Machine Learning Technologies* 2 (2007/2011), S. 37–63

- [RBB⁺13] ROSE, Peter W. ; BI, Chunxiao ; BLUHM, Wolfgang F. ; CHRISTIE, Cole H. ; DIMITROPOULOS, Dimitris ; DUTTA, Shuchismita ; GREEN, Rachel K. ; GOODSSELL, David S. ; PRLIC, Andreas ; QUESADA, Martha ; QUINN, Gregory B. ; RAMOS, Alexander G. ; WESTBROOK, John D. ; YOUNG, Jasmine ; ZARDECKI, Christine ; BERMAN, Helen M. ; BOURNE, Philip E.: The RCSB Protein Data Bank: new resources for research and education. In: *Nucleic Acids Res* 41 (2013), Jan, Nr. D1, D475–D482. <http://dx.doi.org/10.1093/nar/gks1200>. – DOI 10.1093/nar/gks1200
- [RCF⁺01] ROUHANI, S. ; CARTAILLER, J. P. ; FACCIOTTI, M. T. ; WALIAN, P. ; NEEDLEMAN, R. ; LANYI, J. K. ; GLAESER, R. M. ; LUECKE, H.: Crystal structure of the D85S mutant of bacteriorhodopsin: model of an O-like photocycle intermediate. In: *J Mol Biol* 313 (2001), Oct, Nr. 3, 615–628. <http://dx.doi.org/10.1006/jmbi.2001.5066>. – DOI 10.1006/jmbi.2001.5066
- [RRC⁺01] REN, G. ; REDDY, V. S. ; CHENG, A. ; MELNYK, P. ; MITRA, A. K.: Visualization of a water-selective pore by electron crystallography in vitreous ice. In: *Proc Natl Acad Sci U S A* 98 (2001), Feb, Nr. 4, 1398–1403. <http://dx.doi.org/10.1073/pnas.041489198>. – DOI 10.1073/pnas.041489198
- [SBL03] SCHOBERT, Brigitte ; BROWN, Leonid S. ; LANYI, Janos K.: Crystallographic structures of the M and N intermediates of bacteriorhodopsin: assembly of a hydrogen-bonded chain of water molecules between Asp-96 and the retinal Schiff base. In: *J Mol Biol* 330 (2003), Jul, Nr. 3, S. 553–570
- [SBL⁺08] SAPRA, K T. ; BALASUBRAMANIAN, G P. ; LABUDDE, Dirk ; BOWIE, James U. ; MULLER, Daniel J.: Point mutations in membrane proteins reshape energy landscape and populate different unfolding pathways. In: *J Mol Biol* 376 (2008), Feb, Nr. 4, 1076–1090. <http://dx.doi.org/10.1016/j.jmb.2007.12.027>. – DOI 10.1016/j.jmb.2007.12.027
- [Sch10] SCHRÖDINGER, LLC: *The PyMOL Molecular Graphics System, Version 1.3r1*. August 2010
- [SKO57] SKOU, J. C.: The influence of some cations on an adenosine triphosphatase from peripheral nerves. In: *Biochim Biophys Acta* 23 (1957), Feb, Nr. 2, S. 394–401
- [TPF⁺06] TANUJ SAPRA, K. ; PARK, Paul S-H. ; FILIPEK, Slawomir ; ENGEL, Andreas ; MÜLLER, Daniel J. ; PALCZEWSKI, Krzysztof: Detecting molecular interactions that stabilize native bovine rhodopsin. In: *J Mol Biol* 358 (2006), Apr, Nr. 1, 255–269. <http://dx.doi.org/10.1016/j.jmb.2006.02.008>. – DOI 10.1016/j.jmb.2006.02.008
- [VST⁺04] VOGELY, Lutz ; SINESHCHEKOV, Oleg A. ; TRIVEDI, Vishwa D. ; SASAKI, Jun ; SPUDICH, John L. ; LUECKE, Hartmut: Anabaena sensory rhodopsin: a photochromic color sensor at 2.0 Å. In: *Science* 306 (2004), Nov, Nr. 5700, 1390–1393. <http://dx.doi.org/10.1126/science.1103943>. – DOI 10.1126/science.1103943
- [YFY⁺04a] YOHANNAN, Sarah ; FAHAM, Salem ; YANG, Duan ; GROSFELD, David ; CHAMBERLAIN, Aaron K. ; BOWIE, James U.: A C alpha-H...O hydrogen bond in a membrane protein is not stabilizing. In: *J Am Chem Soc* 126 (2004), Mar, Nr. 8, 2284–2285. <http://dx.doi.org/10.1021/ja0317574>. – DOI 10.1021/ja0317574
- [YFY⁺04b] YOHANNAN, Sarah ; FAHAM, Salem ; YANG, Duan ; WHITELEGGE, Julian P. ; BOWIE, James U.: The evolution of transmembrane helix kinks and the structural diversity of G

- protein-coupled receptors. In: *Proc Natl Acad Sci U S A* 101 (2004), Jan, Nr. 4, 959–963. <http://dx.doi.org/10.1073/pnas.0306077101>. – DOI 10.1073/pnas.0306077101
- [YG04] YE, Yuzhen ; GODZIK, Adam: Database searching by flexible protein structure alignment. In: *Protein Sci* 13 (2004), Jul, Nr. 7, 1841–1850. <http://dx.doi.org/10.1110/ps.03602304>. – DOI 10.1110/ps.03602304

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, dass ich meine Arbeit selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und die Arbeit noch nicht anderweitig für Prüfungszwecke vorgelegt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Mittweida, 25. August 2013